IOWA STATE UNIVERSITY
Digital Repository

Graduate Theses and Dissertations

Iowa State University Capstones, Theses and Dissertations

2017

Measurement error modeling of physical activity data

Daniel Ries
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

 Part of the Statistics and Probability Commons

Recommended Citation
Ries, Daniel, "Measurement error modeling of physical activity data" (2017). *Graduate Theses and Dissertations*. 16449.
https://lib.dr.iastate.edu/etd/16449

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Measurement error modeling of physical activity data

by

Daniel Ries

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Alicia Carriquiry, Major Professor
Jarad Niemi
Daniel Nordman
Robin Shook
Lily Wang

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2017

Copyright © Daniel Ries, 2017. All rights reserved.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

ACKNOWLEDGMENTS

I would like to thank my advisor, Professor Alicia Carriquiry for her guidance and support on this dissertation. I am grateful for her enthusiasm and her willingness to work on topics that hold special interest to me. I thank Professors Jarad Niemi, Dan Nordman, and Lily Wang for serving on my committee as well as being great course instructors. I also thank Professor Robin Shook for being on my committee and introducing me to the interesting field of physical activiy and energy balance research. Lastly, I thank Nick Berry, Anand Dixit, and Nehemias Ulloa for their help in keeping graduate school enjoyable.

ABSTRACT

Physical activity is an important component to a healthy lifestyle. However, it is difficult to create physical activity recommendations because it is difficult to track people's physical activity throughout time. This makes it difficult to assess whether people adhere to recommendations or not. Additionally, lack of measurement error-free data makes it difficult to understand the relationships between physical activity and health outcomes.

We address these three issues in this dissertation. First, we construct a flexible measurement error model that uses free knot splines to model the relationship between less expensive measurements and the truth. Because the truth is latent, we model it with a Dirichlet Process mixture prior. We give a calibration algorithm which can be used to eliminate biases in future measurements once this model is fit. Second, we develop a model that allows us to estimate the proportion of people in adherence to the Physical Activity Guidelines (PAG). Additionally, we estimate the entire distribution of usual activity levels which allows us to understand differences among different demographic groups. We find large differences in results using the Physical Activity Measurement Survey (PAMS) from Iowa compared to the 2003-2006 National Health and Nutrition Examination Survey (NHANES), a nationally representative survey. Finally, we construct a model that looks at the relationship between minutes in moderate to vigorous physical activity (MVPA) and Metabolic Syndrome (MetS) and its risk factor components. This model accounts for measurement error as well as the complex data structure. Using the 2003-2006 NHANES data, we find the probability of being diagnosed with MetS goes from 40% to 23% when one participates in 20 minutes of MVPA compared to 0 minutes.

CHAPTER 1. INTRODUCTION

A healthy lifestyle requires an adequate amount of physical activity. Lack of physical activity has been linked to cardiovascular disease (ODPHP (2017), Warburton et al. (2006),Reiner et al. (2013)), diabetes (ODPHP (2017), Warburton et al. (2006), Reiner et al. (2013)), cancer (ODPHP (2017), Warburton et al. (2006)), hypertension (Warburton et al. (2006)), osteoporosis (ODPHP (2017), Warburton et al. (2006)), depression (ODPHP (2017), Warburton et al. (2006)), obesity (Warburton et al. (2006), Reiner et al. (2013)), and Alzheimer's (Reiner et al. (2013)). Tucker et al. (2016), Sisson et al. (2010), and Camhi et al. (2011) all found strong relationships between lack of physical activity and Metabolic Syndrome (MetS). Metabolic syndrome is characterized by high levels of cholesterol, blood pressure, waist circumference, among others and MetS classifies individuals for being high risk for cardiovascular disease and type 2 diabetes. The current Physical Activity Guidelines (PAG) suggest adults should participate in at least 150 minutes each week of moderate-intensity activity, or at least 75 minutes in vigorous-intensity activity, or in some equivalent combination of moderate to vigorous physical activity (MVPA) to achieve adequate health benefits. The PAG also mention that additional health benefits occur when adults exercise beyond the minimum guidelines. In this chapter, we introduce some of the relevant areas of research in the analysis of physical activity data literature, and motivate our ideas to help improve these areas.

1.1 Monitoring Physical Activity

Because physical activity is so important to a healthy lifestyle, there needs to be accurate and precise ways to measure how much activity people participate in or how many calories people burn during a given time period. There are various ways to measure how much and how intense physical activity people participate in. There are two types of measures we consider: i) energy expenditure

(EE) or the amount of calories burned and ii) minutes or MET-minutes at some threshold of activity level. Although these two types of measurements are related, there are also some important differences. EE for our purposes is the total amount of calories expended during a given time frame (say 24 hours). Therefore, this type of measurement considers the aggregate of all types of activity, not only activities that are above some threshold like MVPA, as recommended by the PAG. The amount of calories expended is the sum of dietary induced thermogensis (DIT), volitional physical activity (PA), resting metabolic rate (RMR), and spontaneous physical activity (SPA) (Thomas et al. (2011a)). DIT is related to the energy used when digesting, metabolizing, and transporting food Thomas et al. (2011a). PA represents all fitness related activities such as walking, running, playing sports, etc. We refer to these as voluntary activity, as one could technically live without burning calories this way. RMR is the energy expended in order to keep the body functioning. SPA is the energy used through changes in posture and fidgeting. Note that all four factors will together account for differences in body size, gender, genetics, etc. The other type of physical activity measurement is only looking at the PA component of the above factors. For this we are interested only in how much physical activity does someone participate in at a given level, like MVPA.

1.1.1 Physical activity and energy expenditure measurements

Direct and indirect calorimetry are two methods of obtaining accurate estimates of EE (Levin (2005)). Direct calorimetry measures the rate of heat loss from an individual. This requires an individual to be in a special room for the measurements, so this results in a non-free living state. These special rooms are also costly to build and maintain. Although they may not be realistic measures for EE for a typical day, they can be used to understand how many calories are burned by running a mile at a specific pace, for instance. Indirect calorimetry measure oxygen consumption or carbon dioxide production and use these to calculate EE using formulas. This method is less restrictive than direct calorimetry because it doesn't require a special room, but rather a mask. Again, this would be not be reasonable to measure free-living individuals' EE.

Doubly labeled water (DLW) has been called the gold standard for estimating energy expenditure when individuals are weight stable (Lagerros and Lagiou (2007), Bouten et al. (1996), Thomas et al. (2011b), Hall and Chow (2011), Gilmore et al. (2014), Sanghvi et al. (2015), Racette et al. (2011)). Participants give a baseline urine sample, then drink a small amount of water that contains an irregular, but traceable isotope of hydrogen and oxygen. Two weeks later a urine sample is taken and the concentrations are measured to get an estimate of energy expenditure that is estimated based on washout kinetics Thomas et al. (2011a). DLW is expensive and requires trained personnel to administer and analyze, therefore it is not practical to use DLW in large studies (Hall (2014), Thomas et al. (2011a)). Like direct and indirect calorimetry, DLW is an accurate measure of calories burned, except DLW can be effective in the free-living state since individuals do not need to be in a certain room or connected to a machine. However, DLW cannot give us information on how much time was spent in MVPA, for instance because it gives an aggregate estimate of EE.

Consumer based physical activity monitors with accelerometers have become popular in recent years. Products such as Fitbit, Nike Fuel Band, Jawbone UP band, and Basis B1 Band that measure steps taken, calories burned, heart rate, sleep, GPS, etc. are available to the general public at reasonable cost. They have become more popular in research for measuring EE (Welk et al. (2004)), and their reliability has been confirmed (Bassett (2000), Drenowatz and Eisenmann (2011), Johannsen and et al. (2010), Dannecker (2013),Jung-Min et al. (2014)). The results from these studies are promising, in that physical activity monitors seem to give a relatively inexpensive, accurate, and objective measure of EE at the individual level. These devices will estimate (often with a proprietary algorithms) EE for a given day and also intensity levels and durations as well. This makes these devices potentially useful for measuring how many minutes of MVPA someone participated on a given day, for example.

There are also several "research-grade" wearables on the market. Johannsen and et al. (2010) shows the SenseWear Pro3 armband (SWA) and SenseWear Mini (mini) are relatively good measures of energy expenditure. The SWA has a biaxial accelerometer while the mini has a triaxial accelerometer. Some of what makes these stand apart from popular alternatives like Fitbit, is

they also have sensors to measure heat flux, temperature, and galvanic skin response. One of the concerns with these wearables is assessing compliance. It is often difficult to tell if participants have followed the researcher's directions for using the device. The SWA and mini do not have this problem as they provide estimates of wear time directly (Johannsen and et al. (2010)). Actigraph is another largely used and studied wearable. Actigraphs were used for physical activity assessment part of the National Health and Nutrition Examination Survey (NHANES) in 2003-2006. Unlike the SWA, the Actigraph reports "counts" every minute to denote the amount of movement during that minute, which allows researchers to analyze the raw data instead of processed data under proprietary algorithms like the SWA.

Self reports are also commonly used to measure EE largely due to their ease and cost. They have the benefit of providing more details in terms of type of exercise as well as duration, intensity, context, etc. that an activity monitor would provide. Like with EI, we have the same issue with reporting bias and are not an accurate measure of EE (Neilson et al. (2008)). Respondents typically overestimate their activity and its intensity. However, this inexpensive method can provide a rough activity level estimate and allow one to categorize participants by an approximate activity level if desired.

1.1.2 Energy balance principle and body composition measurements

Physical activity and EE measurements are related to measurements of changes in body composition due to the energy balance principle:

$$\Delta \text{ES} = \text{EI - EE}. \tag{1.1}$$

Therefore we can plausibly gain more information if we collect data on both EE and body composition. If we collect data on both for individuals, we can then jointly model the measurements. This principle is based on the *first law of thermodynamics*, that is, energy can neither be created or destroyed. Energy storage (ES) is composed of fat mass (FM) and fat free mass (FFM). FFM includes muscle tissue, organs, skin, and bones. The ratio of FM and FFM are descriptions of body

composition. ΔES then refers to how one's body composition changes in terms of FM and FFM over some period of time.). Many models have been discussed in the epidemiology and nutrition literature using the energy balance principle to calculate EI using measurements of EE and ΔES (Hall (2010), Thomas et al. (2011a), Thomas et al. (2010), Sanghvi et al. (2015), Hall and Chow (2011), Racette et al. (2011)).

To assess body composition, we use fat mass (FM) and fat free mass (FFM) and assume that total body mass is given by: FM + FFM. The validated equation (Thomas et al. (2012)) below (1.2) shows how we can further break down equation (1.1):

$$\Delta \text{ES} = C_{FFM} \frac{\Delta \text{FFM}}{\Delta \text{t}} + C_{FM} \frac{\Delta \text{FM}}{\Delta \text{t}}, \tag{1.2}$$

where $C_{FFM} = 1020$ and $C_{FM} = 9500$ represent the estimated energy densities of FFM and FM respectively. These estimates are also prone to measurement error, as in a different study by the same author, (Thomas et al. (2010)), the values that were used were $C_{FFM} = 1100$ and $C_{FM} = 9300$.

Setting aside the issue of measuring the energy densities of FM and FFM, we are left with the question of measuring the change in FM and FFM over a specified given of time. Dual-energy X-ray absorptiometry (DXA) is considered the gold standard to measure fat mass (FM) and fat free mass (FFM) (Thomas et al. (2011a),Thomas et al. (2011b),Sanghvi et al. (2015),Gilmore et al. (2014)). DXA scans are typically used to measure bone mineral density, often to diagnose or track osteoporosis, but it can also be to measure body composition. As one might imagine, DXA scans are expensive due to the specialized equipment and personnel needed to obtain the measurement. Hind et al. (2011) show the precision of the new GE Lunar iDXA machine. Coefficient of variation for both FM and FFM were under 1%.

Air displacement plethysmography (ADP) is another approach to measure body composition. ADP measures body volume through Boyle's Law, which describes the inverse relationship between volume and pressure Baracos (2011). After an initial up front cost, ADP machines can be used

quite easily and inexpensively as no highly trained personnel are needed. Measurements can be obtained quickly, and the approach appears to be applicable in populations including the obese and the elderly. Bod Pod is one of the well known measurement tools that performs this.

Bioimpedance spectroscopy (BIS) also provides a relatively accurate and inexpensive tool to measure body composition. BIS instruments have sensors that are placed on the person's extremities (or a subset of them). The tool sends a shock at one end and measures the resistance to reach the other end(s). The results are dependent upon body water levels, which can then be a proxy for FM versus FFM . There are various different levels of quality for these machines. Inexpensive and easy to use versions of BIS tools that are not particularly accurate can be purchased in any nutrition store. Much more expensive, research quality devices are also available. One of the limitations of this approach is its dependence on hydration levels. Since this method provides a measure of body composition that is based on how a shock passes through the body as a function of body water, the person's hydration status can at the time of measurement heavily affect the results.

Finally, calipers are one of the cheapest and simplest tools to measure body composition. The caliper method takes measurements of skinfolds at certain locations to estimate % body fat (which can be used to calculate FM and FFM). The tool itself costs less than $100, and someone can be quickly trained to use a caliper. This type of measurement is subject to significant error (Reilly et al. (1995), Wells et al. (1999)).

1.2 Physical Activity Guidelines

As we've discussed, physical activity is important to our livelihood. The Office of Disease Prevention and Health Promotion (ODPHP) agrees, which is why they have the PAG, and we will focus only on the adult guidelines. As we already mentioned, the guidelines state that adults should participate in at least 150 minutes each week of moderate-intensity activity, or at least 75 minutes in vigorous-intensity activity, or in some equivalent combination of moderate to vigorous physical activity (MVPA), and all these minutes should occur in at least 10 minute bouts of activity. Additionally, adults should do muscle-strengthening activities at least two twice per week targeting

all the major muscle groups. The main reason for the PAG is regular physical activity over the long term can give long term health benefits. The ODPHP website[1] states that the PAG are an "essential resource for health professionals and policymakers" and that the PAG "serves as the primary, authoritative voice of the federal government for evidence-based guidance on physical activity, fitness, and health for Americans." Because of the significance of the PAG to the federal government, it is then important to understand how many Americans adhere to these guidelines. Moreoever, it would be even more useful to understand which demographic groups adhere more or less to these guidelines. The ODPHP website states that fewer than 1 in 4 Americans currently meet the recommended levels of physical activity, but no citation is given for that number. There have been a number of papers that have tried to address the issue of adherence to the PAG. Butcher et al. (2008) found that about 40% of adolescent females and 57% of adolescent males met the physical activity recommendations at the time (pre-PAG), but this was based on a phone survey which includes considerable bias and measurement error. Zhao et al. (2008) also looked at compliance rates to the physical activity recommendations pre-PAG, except for those with and without diabetes. They concluded that compliance rates were about 40% for those with diabetes and about 50% for those without diabetes. Again, the physical activity data was collected via survey. Tucker et al. (2011) used 2005-2006 NHANES accelerometry data and concluded that about 10% of adults would have been in compliance at the time; self-report from NHANES Tucker et al. (2011) found compliance rates to be around 60%, showing a large disparity between the objective and subjective measures. Smith et al. (2016) looked at adherence for African-American breast cancer survivors, and found that "only" 54% reported meeting current PAG. Although the compliance rate estimates are fairly similar for those using self-report data, it is unlikely these are a good representation of true compliance rates due to the biases involved in self-reports. The paper by Tucker used objective measures of physical activity, but there was no mention or consideration of measurement error.

[1] https://health.gov/paguidelines/second-edition/

1.3 The Benefits of Physical Activity

In the opening paragraph, we cited some of the many recorded benefits of physical activity. One area that has a lot of literature is the relationships between physical activity and various MetS risk factors. According to the American Heart Association, MetS affects 34% of American adults which is largely why it has become such a large research topic. Many of the papers focusing on this issue use logistic regression to look at the probability of having MetS for a given level of usual activity (see for example, Tucker et al. (2016) or Sisson et al. (2010)). The results consistently say that those who exercise less have a higher probability for MetS. Recall someone is defined as having MetS if she has a combination of high blood pressure, high blood sugar, high body fat-especially around the waist, and high (low) LDL (HDL) and triglyceride levels. There are fairly well established cut offs for what constitutes "high" levels for each of the previous measures (Tucker et al. (2016)). And a common classifier is if one has at least three measures that are "high", then he is said to have MetS. Therefore much of the current research looks at whether one has MetS or not and physical activity levels, and ignores the fact that the data is continuous for all the measurements and instead uses cutoff points. As a result, there is a loss of information in terms of quantitatively analyzing each risk factor, as well as incorporating the dependencies that exist between these risk factors. What one loses by doing this is understanding exactly how much improvement can be made under a specific exercise regime, as a whole as well as for each individual risk factor. Additionally, none of these studies consider that the data collected on physical activity contain measurement error, which is known to cause bias in parameter estimates (Fuller (1987)).

1.4 Overview of Dissertation

In this dissertation, we propose measurement error models to address some of the current open issues in analyzing physical activity data. We take a Bayesian approach to these models to provide an alternative to the more common frequentist measurement error models as well as to provide models that are more intuitive in development, estimation, and interpretation. In Chapter 2, we develop methodology to assess the biases and calibrate less expensive measurements by jointly

modeling EE and ΔES gold standard and less expensive measurements. This joint model allows us to gain information from both physical activity side and body composition size of measurements, which are related through the energy balance principle 1.1. We purposely develop a flexible model that incorporates semi-parametric regression including free-knot splines to model the biases of less expensive measurements, and we use Dirichlet Process Mixture Models to model latent variables to provide an alternative to the standard Normal model. We give in detail a Reversible Jump Markov Chain Monte Carlo algorithm to estimate this semi-parametric model. Finally, we provide a method to do calibration of less expensive measurements when all parameters have been estimated. We do a simulation study to show results of this model under realistic data. In Chapter 3, we present a model to assess compliance to the 2008 PAG using data from the Physical Activity Measurement Survey and NHANES. This model not only addresses the issues of how many people adhere to the PAG, but also looks at who. Additionally, we calculate distributions of usual physical activity levels for various populations. In Chapter 4 we model MetS risk factors as a function of measurement error corrected minutes of MVPA. Instead of classifying people as having MetS or not, we model the individual risk factor levels for each individual jointly on the continuous scale using both linear and nonlinear functions. Instead of assuming normality for the regression errors, we use a mixture of multivariate normal distributions because many of the risk factors have skewed distributions. For this model we use data from 2003-2006 NHANES. In Chapter 5, we make final conclusions and discuss some areas for future research.

CHAPTER 2. MODELING ENERGY BALANCE WHILE CORRECTING FOR MEASUREMENT ERROR WITH FREE KNOT SPLINES VIA REVERSIBLE JUMP MCMC

2.1 Introduction

Accurately measuring energy balance in free-living individuals is challenging, even in small studies. Yet to design effective public health policies and interventions, it would be valuable to be able to assess energy balance in nationwide surveys such as National Health and Nutrition Examination Survey (NHANES). Clearly, instruments such as doubly labeled water (DLW) and dual-energy X-ray absorptiometry (DXA) are too costly and burdensome to administer in large groups. Therefore, a question of interest is whether measurements of energy balance obtained from self-report instruments or even from objective measuring tools such as the Sensewear Armband, which are much less costly to apply, can be calibrated. In this chapter, we explore the association between measurements obtained from accurate instruments and those obtained from noisy instruments which can be administered to large groups. We are interested in formulating a model for energy balance by using energy expenditure (EE) and changes in energy stores (ΔES) while accounting for dependence between the two and measurement error. Widely accepted gold standard measurements exist for both EE (DLW) (Lagerros and Lagiou (2007), Bouten et al. (1996), Thomas et al. (2011b), Hall and Chow (2011), Gilmore et al. (2014), Sanghvi et al. (2015), Racette et al. (2011)) and ΔES (DXA) (Thomas et al. (2011b), Sanghvi et al. (2015), Gilmore et al. (2014)). Unfortunately, these instruments are expensive and burdensome. There are alternative approaches to quantify both EE and ΔES that while less expensive and easier to administer, are subject to bias and other errors. Our goal is to model energy balance by using both gold standard and inexpensive instruments with the end goal of evaluating the error present in the measurements and ultimately calibrating the

inexpensive instruments, so in future studies, researchers can calibrate their measurements of EE and/or ΔES if they are not using a gold standard.

In this chapter we adopt a Bayesian semi-parametric approach. We make distributional assumptions about error terms, but we try to be flexible when modeling the true relationship between inexpensive measurements and the truth. We propose using free knot splines to model the relationship between the inexpensive measurements and the truth and we build a Reversible Jump MCMC algorithm to do so. This analysis requires that we obtain replicate observations for the gold standard and inexpensive measurements on each participant in the study, or at least on a random sub-sample. This is expensive and time consuming in terms of data collection, but the dataset permits addressing uncertainty at different levels. We use data from the Energy Balance Study to assist in constructing a model (Hand et al. (2013)).

2.2 Energy Balance Study Data

The Energy Balance Study (EBS)[1] (Hand et al. (2013)) was conducted in 2011 and 2012 at the University of South Carolina. Over the span of a year, measurements of EI, EE, and ΔES were obtained from 430 men and women between the ages of 21 and 35 and with BMIs ranging from 20 to 35. Study participants received 5 DXA scans, one at baseline, and one every 3 months thereafter. To measure EI, multiple 24 hour recalls were administered over the phone, a set at baseline and then a set every 3 months up to a year. Average values of total calories consumed per day were recorded every 3 months. EE was measured with SenseWear Armbands over ten day periods, at baseline, and then every 3 months there after for up to a year. During the assessment period every three months, other measures such as weight, BMI, hip and waist circumference were obtained as well. On a subset of the participants ($n = 191$), DLW was administered after their 12 months of the study was complete. At the beginning and end of the two week period during which EE is assessed via DLW, DXA scans were obtained as well. All EE and EI measurements were recorded in kcals and were averaged over the assessment period to a single, daily average. DXA

[1]Funding for this project was provided through an unrestricted grant from The Coca-Cola Company. The authors wish to thank the study participants and the Energy Balance Study team.

gives measurements of FM and FFM in kg at a certain point in time. Note, we do not technically have replicates on individuals since measurements are taken every three months. However, we will use these data for explortatory purposes.

2.2.1 Calculation of ΔES

In the energy balance equation,

$$\Delta ES = EI - EE, \tag{2.1}$$

ΔES is expressed in kcals, and can be positive or negative. To convert DXA measurements of fat mass and fat free mass to kcals, we use equation (2.2). Because we assume that energy stores are characterized only as either fat mass (FM) or fat free mass (FFM), this equation provides an exact answer if we know the values of C_{FM} and C_{FFM}. We let $C_{FM} = 9500$ and $C_{FFM} = 1100$ like in Thomas et al. (2011b), recognizing that a single value does not account for biological variation. We divide these by the change in time (14 days ± 3 days) and multiply by C_{FM} and C_{FFM} to get ΔES in kcals. For each individual, we compute

$$\Delta ES = C_{FM}\frac{\Delta FM}{\Delta T} + C_{FFM}\frac{\Delta FFM}{\Delta T}. \tag{2.2}$$

2.2.2 Characteristics of EBS data

The goal in this exploratory exercise is to determine the general form and distribution of the variables as well as check some common measurement error model assumptions. We also want to look at the underlying relationships between measurements and demographic factors. In Figure 2.1 we plot the distribution of EE (left) and ΔES (right) as measured by gold standard instruments, by gender. For these figures we use the sub-sample of participants who received DLW. We see that the observed distribution of EE by DLW is slightly skewed right with males having greater EE overall. The change in body composition ΔES by DXA is fairly symmetric and there are no obvious differences between genders.

In Figure 2.2 we show diagnostic plots to check whether we can plausibly assume constant variance for these specific measures of EE and ΔES, respectively. Both of these plots show the mean based on replicates for an individual on the x axis and that same individual's standard deviation based on replicates on the y axis. The left plot is for EE as measured by the Sensewear Armband and the right plot is for ΔES as measured by DXA. Looking at the EE plot, we see some signs of non-constant variance, with variability increasing with EE. The ΔES plot does not show much in terms of non-constant variability.

Figure 2.3 shows normal quantile plots for DLW measured EE, by gender, females on left and males on right. Note that this is not the usual measurement error diagnostic plot suggested in Caroll et al. (2006) since it is based on actual EE values and not differenced values, however we don't have replicates for DLW measurements. Because DLW is the gold standard for EE we have at least some empirical evidence of normality for these measurement errors.

Figure 2.4 shows normal quantile plots for differenced DXA measurements of ΔES, by gender, females on left and males on right. We calculated the differenced ΔES for 6 month DXA to 9 month DXA and 9 month DXA to 12 month DXA and computed the difference. Both plots look similar and show small signs of non normality in the tails. Figure 2.5 shows normal quantile plots for differenced armband measured EE, by gender. We calculated the differenced EE by taking the 12 month measurement minus the 9 month measurement. Again we show females on left and males on right. There do not appear to be any gross departures from normality.

We checked for evidence of systematic bias in the inexpensive measurement of EE, which based on the EBS data is the Sensewear Armband. Using only the subset of individuals who received DLW at the end of the study, we fit a multiple regression using DLW-measured EE, age, BMI, and gender as covariates. Regression results are given in Table (2.1). Not surprisingly, the coefficient on DLW-measured EE was significant. The coefficients on gender and BMI also were highly significant as indicated by small p-values and large effect sizes. Age was only weakly significant, but again the EBS study cohort only included a small age range and this finding could change with a larger population. We choose to include age as well as a linear bias term in our model. We note that

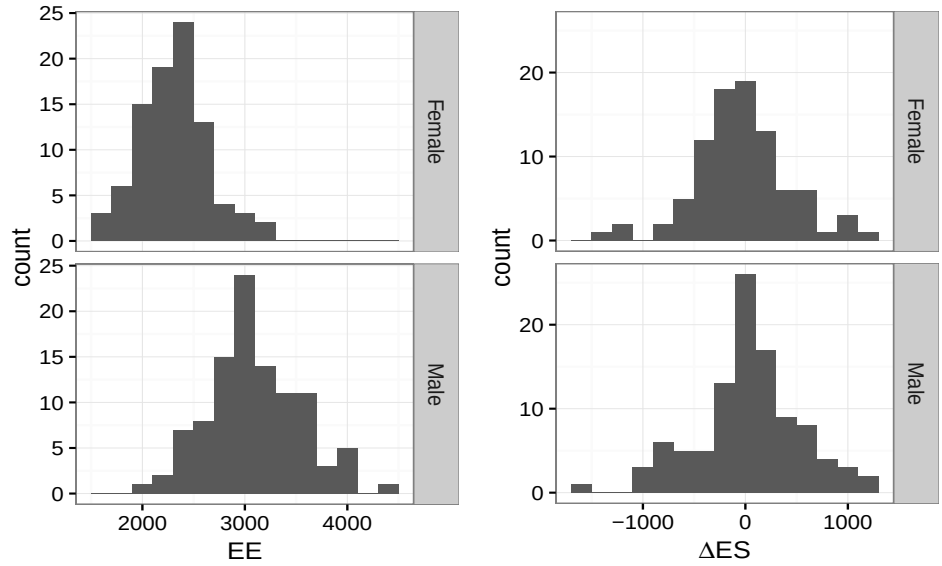

Figure 2.1: Distribution of EE (left) measured by DLW and ΔES measured by DXA (right). Females on top, Males on bottom. Includes data from EBS from individuals who received DLW measurements.

we are fitting this regression using DLW measured EE and not *true, usual* EE as specified in (2.4),(2.5). This simple regression model provides evidence that systematic biases do exist in inexpensive measurements of EE, and we assume that the same holds true for inexpensive ΔES measurements.

2.3 Methodology

2.3.1 Notation

We denote observed average daily EE measured via DLW for subject i over time period j by W_{ij}^{EE}, and observed average daily ΔES measured via DXA for subject i over time period j by $W_{ij}^{\Delta ES}$. A positive value for ΔES indicates that more calories were consumed than expended. We compute daily values of EE for a person by averaging the total EE for that person obtained by DLW, because DLW gives an estimate of EE over the course of approximately 14 days.

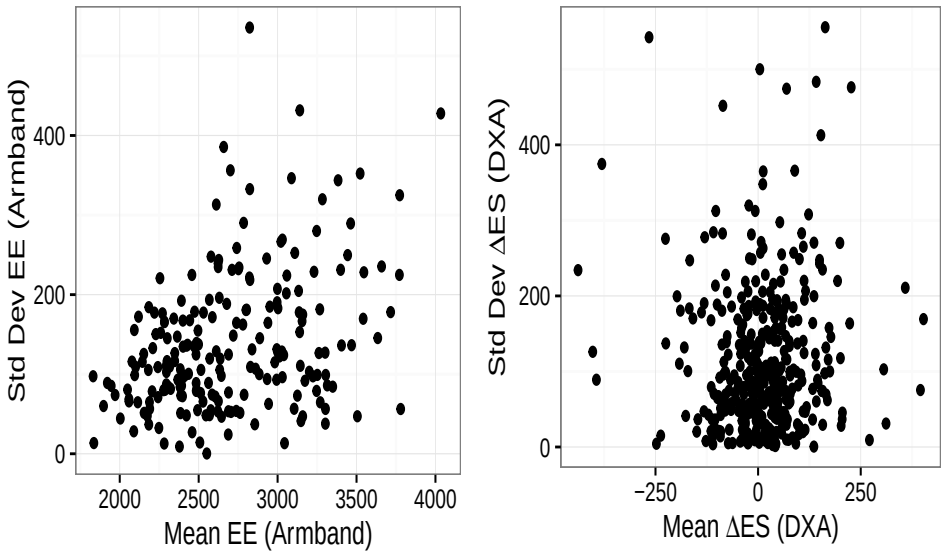

Figure 2.2: Variance Diagnostic Plots. The left shows the mean EE (Armband) for an individual vs his/her standard deviation of EE based on replicates. The right is the same except for ΔES measured by DXA.

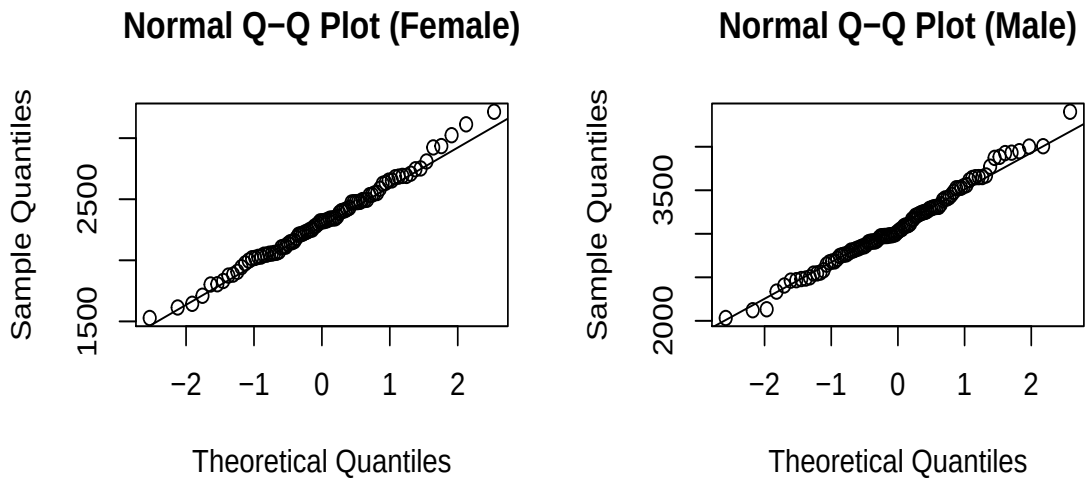

Figure 2.3: Normal Quantile Plots for EE as measured by DLW for Female (left) and Male (right). Data is from those who received DLW in EBS.

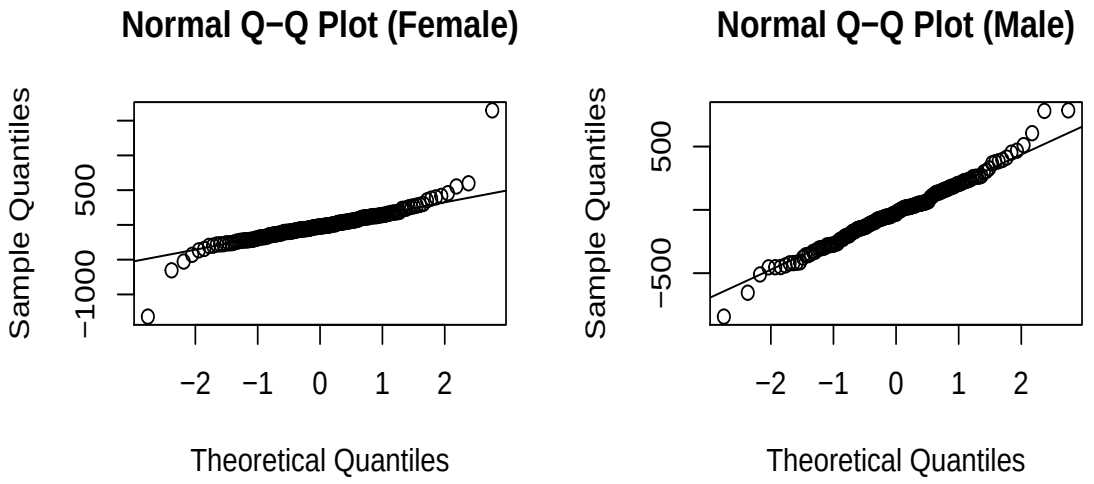

Figure 2.4: Normal Quantile Plots for differenced ΔES ((12m DXA - 9m DXA)-(9m DXA - 6m DXA)) as measured by DXA for Female (left) and Male (right). Data is from those who had DXA scans at 6,9,12 months

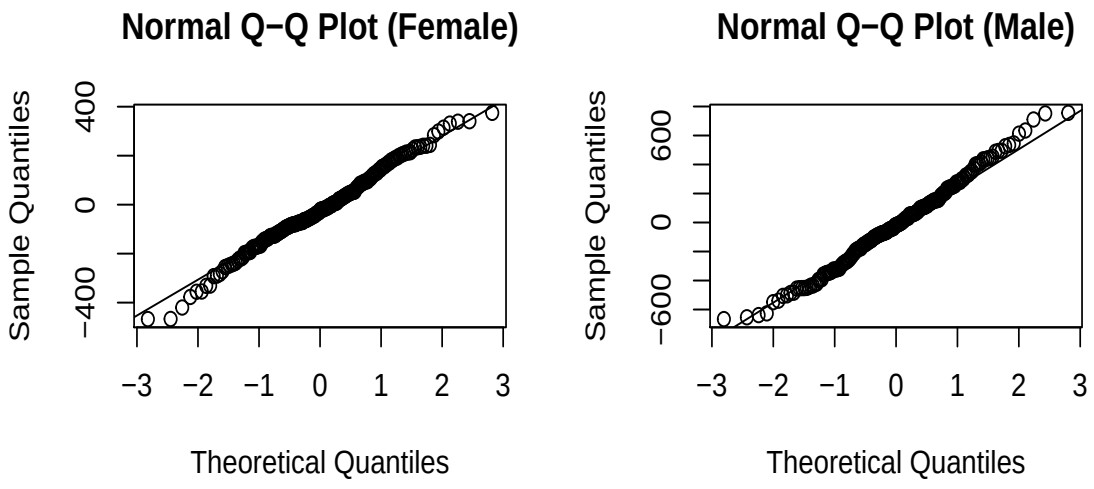

Figure 2.5: Normal Quantile Plots for differenced EE (12m EE - 9m EE) as measured by armbands for Female (left) and Male (right). Based on entire EBS individuals who remained for entire 12 months

Table 2.1: Regression of Sensewear Armband EE on DLW EE, Age, BMI, Gender using EBS data. Results show systematic biases could exist in inexpensive EE measurements.

	Dependent variable:
	Energy_expenditure
TDEE	0.559***
	(0.040)
age	−7.351*
	(4.000)
genderMale	305.583***
	(43.259)
BMI	14.146***
	(3.989)
Constant	878.423***
	(154.113)
Observations	191
R^2	0.830
Adjusted R^2	0.826
Residual Std. Error	206.875 (df = 186)
F Statistic	226.938*** (df = 4; 186)
Note:	*p<0.1; **p<0.05; ***p<0.01

When collecting data on a large population, it is practical to administer less expensive instruments on most of the subjects. However, they result in less accurate measurements. Although there are several inexpensive ways to measure EE and ΔES, we keep the notation general since in any given situation we will refer only to one specific instrument. We denote the observed average daily EE obtained with an inexpensive instrument for subject i over time period j, Y_{ij}^{EE}, and the observed average daily change in energy stores measured by an inexpensive instrument for subject i over time period j, $Y_{ij}^{\Delta ES}$.

Lastly, the values which we cannot observe are the *usual* EE and ΔES for subject i. We define *usual* as a long run average (expected value) of the true EE and ΔES. Let X_i^{EE} represent the *usual* daily EE for subject i and $X_i^{\Delta ES}$ represent the *usual* daily ΔES for subject i. Note that even if we could observe daily EE and daily ΔES for each participant with no error, there is still

within-person variability in these two variables because people change their caloric intake and their physical activity from day to day.

2.3.2 (In)Dependence assumptions

The observed data vector for subject i at time j is $(W_{ij}^{EE}, W_{ij}^{\Delta ES}, Y_{ij}^{EE}, Y_{ij}^{\Delta ES}, Z_i)$ where Z_i is a vector of covariates measured with no error for subject i. We start by assuming independence between individuals. A joint probability model for these observables for an individual i is represented as:

$$f(W_{ij}^{EE}, W_{ij}^{\Delta ES}, Y_{ij}^{EE}, Y_{ij}^{\Delta ES} | Z_i, \boldsymbol{\theta}). \tag{2.3}$$

Several of the variables in the model are conditionally independent. Given the value X_i^{EE} (*usual* daily EE for subject i) and $X_i^{\Delta ES}$ (*usual* daily ΔES for subject i), and covariates Z_i, we assume that:

1. Y_{ij}^{EE} and $Y_{ij}^{\Delta ES}$ are independent of each other, and
2. Y_{ij}^{EE} and $Y_{ij}^{\Delta ES}$ are each independent of both W_{ij}^{EE} and $W_{ij}^{\Delta ES}$,
3. W_{ij}^{EE} and $W_{ij}^{\Delta ES}$ are independent of each other.

Assumption (1) follows because given the true values X and covariates Z, knowing an inexpensive measurement will give us no more information about the inexpensive measurement of the other, so long as it is not self-administered. To justify assumption (2), we note that once we know the truth X, having an unbiased measurement of X will not provide any more information about the inexpensive, biased measurement of X. Assumption (3) follows from a reasoning similar to (1).

2.3.3 Model likelihood

We propose the following models for the W's and Y's:

$$Y_{ij}^{EE} = m_{ee}(X_i^{EE}, Z_i) + \epsilon_{ij}^{EE} \tag{2.4}$$

$$Y_{ij}^{\Delta ES} = m_{es}(X_i^{\Delta ES}, Z_i) + \epsilon_{ij}^{\Delta ES} \tag{2.5}$$

$$W_{ij}^{EE} = X_i^{EE} + \nu_{ij}^{EE} \tag{2.6}$$

$$W_{ij}^{\Delta ES} = X_i^{\Delta ES} + \nu_{ij}^{\Delta ES}, \tag{2.7}$$

where $m_{ee}(\cdot)$ and $m_{es}(\cdot)$ are assumed to be unknown, smooth functions, ν_{ij}^{EE} and $\nu_{ij}^{\Delta ES}$ represent the measurement error for subject i during time period j using DLW and DXA, respectively, and within person variability. For now, we let ϵ_{ij}^{EE} and $\epsilon_{ij}^{\Delta ES}$ denote the measurement error for subject i during time period j using inexpensive measurements of EE and ΔES, respectively, and within person variability. This is a working assumption since the within person variability is in reality not an additive error unless m_{ee} and m_{es} are linear functions. We assume that the expected values of ϵ, ν are 0, which implies that DLW and DXA are unbiased for *usual* daily EE (X_i^{EE}) and ΔES ($X_i^{\Delta ES}$), respectively. The forms of 2.4-2.7 resemble the general structure for Bayesian nonparametric measurement error modeling first proposed by Berry et al. (2002) and later extended by Sarkar et al. (2014b).

Conditional on each subjects' *usual* value X_i, replicate measurements within subject are independent. Independence within individuals is justified because it is conditional *on each individuals usual value*, which means once we know an individual's baseline, we assume that their daily measurements are independent from each other. The likelihood function for individual i can then be written as:

$$
\begin{aligned}
L_i(\boldsymbol{\theta}) &= \prod_{j=1}^{J} f(W_{ij}^{EE}, W_{ij}^{\Delta ES}, Y_{ij}^{EE}, Y_{ij}^{\Delta ES} | Z_i, \boldsymbol{\theta}) \\
&= \int_{\mathcal{X}_{es}} \int_{\mathcal{X}_{ee}} \prod_{j=1}^{J} f(W_{ij}^{EE}, W_{ij}^{\Delta ES}, Y_{ij}^{EE}, Y_{ij}^{\Delta ES}, X_i^{EE}, X_i^{\Delta ES} | Z_i, \boldsymbol{\theta}) dX_i^{EE} dX_i^{\Delta ES} \\
&= \prod_{j=1}^{J} \int_{\mathcal{X}_{es}} \int_{\mathcal{X}_{ee}} f(W_{ij}^{EE} | X_i^{EE}, X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{wee}}) f(W_{ij}^{\Delta ES} | X_i^{EE}, X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{wes}}) \times \\
&\quad f(Y_{ij}^{EE} | X_i^{EE}, X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{yee}}) f(Y_{ij}^{\Delta ES} | X_i^{EE}, X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{yes}}) \times \\
&\quad f(X_i^{EE}, X_i^{\Delta ES} | Z_i, \boldsymbol{\theta_x}) dX_i^{EE} dX_i^{\Delta ES} \\
&= \prod_{j=1}^{J} \int_{\mathcal{X}_{es}} \int_{\mathcal{X}_{ee}} f(W_{ij}^{EE} | X_i^{EE}, Z_i, \boldsymbol{\theta_{wee}}) f(W_{ij}^{\Delta ES} | X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{wes}}) \times \\
&\quad f(Y_{ij}^{EE} | X_i^{EE}, Z_i, \boldsymbol{\theta_{yee}}) f(Y_{ij}^{\Delta ES} | X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{yes}}) \times \\
&\quad f(X_i^{EE}, X_i^{\Delta ES} | Z_i, \boldsymbol{\theta_x}) dX_i^{EE} dX_i^{\Delta ES},
\end{aligned}
\tag{2.8}
$$

where the RHS is expanded in (2.8). Since we assume that individuals are independent, the full likelihood is

$$
L(\boldsymbol{\theta}) = \prod_{i=1}^{n} L_i(\boldsymbol{\theta}). \tag{2.9}
$$

2.3.4 Naïve model

The first model we consider is what we call the naïve model. This model assumes no measurement error in the gold standard instrument, thus DLW and DXA give error-free measurements of X_i^{EE} and $X_i^{\Delta ES}$, respectively. We also assume that the inexpensive measurements Y are linearly related to the *usual* values and to error free covariates. Based on empirical evidence from EBS data, gender, BMI, and age all had some effect on the inexpensive measurement of EE. The naïve model is:

$$Y_{ij}^{EE} = \beta_{0,ee} + \beta_{1,ee} W_{ij}^{EE} + \boldsymbol{\gamma_{ee}} Z_i + \epsilon_{ij}^{EE} \tag{2.10}$$

$$Y_{ij}^{\Delta ES} = \beta_{0,es} + \beta_{1,es} W_{ij}^{\Delta ES} + \boldsymbol{\gamma_{es}} Z_i + \epsilon_{ij}^{\Delta ES}, \tag{2.11}$$

where the $\beta_{1,\cdot}$ terms represents the relationship between inexpensive measurements and the *usual* EE and ΔES and the β_0 terms represent systematic biases. We let $\boldsymbol{\gamma}_{\cdot} = (\gamma_{1,\cdot}, \gamma_{2,\cdot}, \gamma_{3,\cdot})$ and $\gamma_{1,\cdot}$ is the coefficient for gender, $\gamma_{2,\cdot}$ is the coefficient for BMI, and $\gamma_{3,\cdot}$ is the coefficient for age. We take the standard approach and assume that the errors are normally distributed so that:

$$(Y_{ij}^{EE} | W_{ij}^{EE}, Z_i, \boldsymbol{\theta_{yee}}) \overset{ind}{\sim} N(\beta_{0,ee} + \beta_{1,ee} W_{ij}^{EE} + \boldsymbol{\gamma_{ee}} Z_i, \sigma^2_{\epsilon^{EE}}) \tag{2.12}$$

$$(Y_{ij}^{\Delta ES} | W_{ij}^{\Delta ES}, Z_i, \boldsymbol{\theta_{yes}}) \overset{ind}{\sim} N(\beta_{0,es} + \beta_{1,es} W_{ij}^{\Delta ES} + \boldsymbol{\gamma_{es}} Z_i, \sigma^2_{\epsilon^{\Delta ES}}). \tag{2.13}$$

We choose independent priors for all model parameters. Where appropriate, we select priors that are conjugate or conditionally conjugate for ease of implementation but also to permit incorporating weak information through the prior. Priors are shown below:

$$\beta_{0,ee} \sim N(M_{\beta_{0,ee}}, C_{\beta_{0,ee}}) \tag{2.14}$$

$$\beta_{1,ee} \sim N(M_{\beta_{1,ee}}, C_{\beta_{1,ee}}) \tag{2.15}$$

$$\gamma_{g,ee} \sim N(M_{\gamma_{ee}}, C_{\gamma_{ee}}), \ g = 1, 2, 3 \tag{2.16}$$

$$\sigma^2_{\epsilon^{ee}} \sim IG(a_{yee}, b_{yee}) \tag{2.17}$$

$$\beta_{0,es} \sim N(M_{\beta_{0,es}}, C_{\beta_{0,es}}) \tag{2.18}$$

$$\beta_{1,es} \sim N(M_{\beta_{1,es}}, C_{\beta_{1,es}}) \tag{2.19}$$

$$\gamma_{g,es} \sim N(M_{\gamma_{es}}, C_{\gamma_{es}}), \ g = 1, 2, 3 \tag{2.20}$$

$$\sigma^2_{\epsilon^{es}} \sim IG(a_{yes}, b_{yes}). \tag{2.21}$$

IG refers to the inverse gamma distribution, and $\sigma^2_{\epsilon^{ee}}$ and $\sigma^2_{\epsilon^{es}}$ represent the measurement error and within-person variability as measured in the inexpensive EE and ΔES tools, respectively.

2.3.5 Linear measurement error model (LMEM)

The Linear Measurement Error Model (LMEM) model is similar to the naïve model except that we explicitly recognize that W^{EE} and $W^{\Delta ES}$ are contaminated with additive measurement error.

Instead of assuming that the gold standard measurements are error free, we let them follow the form of (2.6),(2.7). We assume that the measurement errors have independent normal distributions. We take the common approach and model $(X_i^{EE}, X_i^{\Delta ES})$ with a bivariate normal distribution. More formally, the model is given by:

$$(Y_{ij}^{EE}|X_i^{EE}, Z_i, \boldsymbol{\theta_{yee}}) \overset{ind}{\sim} N(\beta_{0,ee} + \beta_{1,ee}X_i^{EE} + \boldsymbol{\gamma_{ee}}Z_i, \sigma^2_{\epsilon_{EE}}) \tag{2.22}$$

$$(Y_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{yes}}) \overset{ind}{\sim} N(\beta_{0,es} + \beta_{1,es}X_i^{\Delta ES} + \boldsymbol{\gamma_{es}}Z_i, \sigma^2_{\epsilon_{\Delta ES}}) \tag{2.23}$$

$$(W_{ij}^{EE}|X_i^{EE}, Z_i, \boldsymbol{\theta_{wee}}) \overset{ind}{\sim} N(X_i^{EE}, \sigma^2_{\nu_{EE}}) \tag{2.24}$$

$$(W_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{wes}}) \overset{ind}{\sim} N(X_i^{\Delta ES}, \sigma^2_{\nu_{\Delta ES}}) \tag{2.25}$$

$$(X_i^{EE}, X_i^{\Delta ES}|\theta_X) \overset{ind}{\sim} N\left(\begin{bmatrix}\mu_{EE}\\ \mu_{\Delta ES}\end{bmatrix}, \Sigma_X\right). \tag{2.26}$$

We choose the following prior distributions:

$$\beta_{0,ee} \sim N(M_{\beta_{0,ee}}, C_{\beta_{0,ee}}) \tag{2.27}$$

$$\beta_{1,ee} \sim N(M_{\beta_{1,ee}}, C_{\beta_{1,ee}}) \tag{2.28}$$

$$\gamma_{g,ee} \sim N(M_{\gamma_{ee}}, C_{\gamma_{ee}}),\ g = 1, 2, 3 \tag{2.29}$$

$$\sigma^2_{\epsilon^{ee}} \sim IG(a_{yee}, b_{yee}) \tag{2.30}$$

$$\sigma^2_{\nu^{ee}} \sim IG(a_{wee}, b_{wee}) \tag{2.31}$$

$$\Sigma_X \sim Inv - Wish(\psi, d) \tag{2.32}$$

$$\beta_{0,es} \sim N(M_{\beta_{0,es}}, C_{\beta_{0,es}}) \tag{2.33}$$

$$\beta_{1,es} \sim N(M_{\beta_{1,es}}, C_{\beta_{1,es}}) \tag{2.34}$$

$$\gamma_{g,es} \sim N(M_{\gamma_{es}}, C_{\gamma_{es}}),\ g = 1, 2, 3 \tag{2.35}$$

$$\sigma^2_{\epsilon^{es}} \sim IG(a_{yes}, b_{yes}) \tag{2.36}$$

$$\sigma^2_{\nu^{es}} \sim IG(a_{wes}, b_{wes}) \tag{2.37}$$

$$\mu_{EE}, \mu_{\Delta ES} \sim N(M_\mu, C_\mu) \tag{2.38}$$

2.3.6 Spline measurement error model (SMEMDP)

We extend the LMEM for EE and ΔES in the previous section to include both parametric and non-parametric components. The goal is to be able to have interpretable parameters, but also to have a model that is flexible. Following (2.8), we explain how we model every component to arrive at a joint probability model.

We follow the same construction of the LMEM to model the gold standard measurements as unbiased for *usual* attributes and subject to normally distributed measurement errors:

$$(W_{ij}^{EE}|X_i^{EE}, Z_i, \sigma^2_{\nu^{EE}}) \overset{ind}{\sim} N(X_i^{EE}, \sigma^2_{\nu^{EE}}) \tag{2.39}$$

$$(W_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \sigma^2_{\nu^{\Delta ES}}) \overset{ind}{\sim} N(X_i^{\Delta ES}, \sigma^2_{\nu^{\Delta ES}}). \tag{2.40}$$

We start with the structure in (2.4),(2.5) for Y_{ij}^{EE} and $Y_{ij}^{\Delta ES}$. We have empirical evidence that inexpensive instruments are not only noisy, but biased as well. We wish to understand both the biases as functions of *usual* value and demographic covariates, as well as the measurement error in the instruments themselves. We propose modeling the inexpensive measurements in a semi-parametric regression framework. Specifically, model the functions $m_\cdot(\cdot)$ using free knot cubic B-splines, and model demographic covariates with a linear component. We require monotone functions so we can take inverses for calibration later, but this only requires the spline coefficients to be non-decreasing ie. $\beta_1 \leq \beta_2 \leq ... \leq \beta_k$ (de Boor (1978)) as used in similar applications (Sinha et al. (2010), Leitenstorfer and Tutz (2007), Wang and Small (2015)). Our approach has three benefits. First, the spline is flexible and can pick up an unknown relationship between $X^\cdot$ and the inexpensive measurement of the same, which is important because we never observe the truth and therefore it is difficult to justify a particular functional form of the relationship. Second, the use of free knot splines eliminates the need for us to specify the number and position of the knots. Previous methods using splines in measurement error models choose a "moderately large" number of knots, typically at least 15 (Sinha et al. (2010), Sarkar et al. (2014b), Berry et al. (2002)). We use Reversible Jump MCMC (RJMCMC) to determine the number and position of knots. This means

that we treat the number of knots in each regression equation and their knot locations as random variables. Third, the linear component for the covariates allows for an easy interpretation of the parameters and thus the biases in the instrument. This is convenient because the covariates we use are typically collected in most exercise, diet, or health study and they can help better calibrate the inexpensive instruments. We make a working assumption of constant variance for all measurement errors at this time. Based on the above, the model specification is then:

$$(Y_{ij}^{EE}|X_i^{EE}, Z_i, \boldsymbol{\theta_{yee}}) \overset{ind}{\sim} N(s_{ee}(X_i^{EE}; \boldsymbol{\beta_{ee}}) + \boldsymbol{\gamma_{ee}} Z_i, \sigma^2_{\epsilon^{EE}}) \tag{2.41}$$

$$(Y_{ij}^{\Delta ES}|X_i^{\Delta ES}, Z_i, \boldsymbol{\theta_{yes}}) \overset{ind}{\sim} N(s_{es}(X_i^{\Delta ES}; \boldsymbol{\beta_{\Delta es}}) + \boldsymbol{\gamma_{es}} Z_i, \sigma^2_{\epsilon^{\Delta ES}}) \tag{2.42}$$

$$s_{ee}(X_i^{EE}; \boldsymbol{\beta_{ee}}) = \sum_{i=1}^{k_{ee}+4} b_{i,ee}(\mathbf{X^{EE}})\beta_{i,ee} = B_{ee}(\mathbf{X^{EE}})\boldsymbol{\beta_{ee}} \tag{2.43}$$

$$s_{es}(X_i^{\Delta ES}; \boldsymbol{\beta_{\Delta es}}) = \sum_{i=1}^{k_{es}+4} b_{i,es}(\mathbf{X^{\Delta ES}})\beta_{i,es} = B_{es}(\mathbf{X^{\Delta ES}})\boldsymbol{\beta_{es}}, \tag{2.44}$$

where $B_{ee}()$ and $B_{es}()$ are $n \times (k_{ee} + 4)$ and $n \times (k_{es} + 4)$ B-spline basis matrices that can be constructed using the recursion specified in Takezawa (2006). We let k_{ee} and k_{es} denote the number of knots for the EE and ΔES splines, respectively.

There are many different types of splines, but we picked B-splines because in similar problems (Sinha et al. (2010), Sarkar et al. (2014a), Sarkar et al. (2014b)) it has been shown that they are numerically more stable than P-splines, for example, which can have major effects on outcomes as compared in Sarkar et al. (2014a).

We allow more flexibility in the distribution of the latent variables $(X_i^{EE}, X_i^{\Delta ES})$ by specifying a Dirichlet process mixture prior for them. This allows the data to "speak for themselves" which is ideal when the model includes latent variables. The density of $(X_i^{EE}, X_i^{\Delta ES})$ can then be modeled as an infinite mixture of normals. Then we can model the distribution for the latent variables as:

$$(X_i^{EE}, X_i^{\Delta ES})|\zeta_i = h \overset{iid}{\sim} N\left(\begin{bmatrix}\mu_{ee,h}\\ \mu_{es,h}\end{bmatrix}, \Sigma_h\right) \tag{2.45}$$

$$\zeta_i \overset{iid}{\sim} Cat(H, \boldsymbol{\pi}) \tag{2.46}$$

$$V_h \sim \text{Beta}(1, \alpha) \tag{2.47}$$

$$V_H = 1 \tag{2.48}$$

$$\pi_h = V_h \prod_{\ell<h}(1 - V_\ell), \tag{2.49}$$

where α helps control how many components of the infinite mixture are used. We choose to set α to 1. The parameter ζ_i takes value for which group observation i came. $Cat(H, \boldsymbol{\pi})$ is a categorical random variable such that $P(\zeta_i = h) = \pi_h, h \leq H$. In any given problem, we can select H such that $\sum_{h=1}^{H} \pi_h < \epsilon$ for some $\epsilon > 0$ (Gelman et al. (2014), pg. 552) . To help simplify the construction of the sampler, we choose conditionally conjugate priors when available. We also let all priors be independent of each other. Specifically,

$$\sigma^2_{\epsilon^{ee}} \sim IG(a_{yee}, b_{yee}) \tag{2.50}$$

$$\sigma^2_{\nu^{ee}} \sim IG(a_{yes}, b_{yes}) \tag{2.51}$$

$$k_{ee} \sim Poi(\lambda_{ee}) \tag{2.52}$$

$$r_{ee} \sim DisUnif(X_1^{EE}, ..., X_n^{EE}) \tag{2.53}$$

$$\Sigma_h \sim Inv - Wish(\psi, d), h = 1, ...H \tag{2.54}$$

$$\sigma^2_{\epsilon^{es}} \sim IG(a_{wee}, b_{wee}) \tag{2.55}$$

$$\sigma^2_{\nu^{es}} \sim IG(a_{wes}, b_{wes}) \tag{2.56}$$

$$k_{es} \sim Poi(\lambda_{es}) \tag{2.57}$$

$$r_{es} \sim DisUnif(X_1^{\Delta ES}, ..., X_n^{\Delta ES}) \tag{2.58}$$

$$\mu_{EE,h}, \mu_{\Delta ES,h} \sim N(M_\mu, C_\mu), h = 1, ...H. \tag{2.59}$$

Although we do not know the true form of the association between the noisy measurements and the usual values, we do not anticipate it to be highly complex, so we would like to use as few knots as necessary. We use r_{ee} and r_{es} to denote the knot locations. Our discrete uniform prior on these, means that knots can only occur at the latent values of $(X_i^{EE}, X_i^{\Delta ES})$. This was done largely for computational convenience; we could have assigned a continuous prior for the knot locations, but we do not believe this will adversely affect estimation because the latent $(X_i^{EE}, X_i^{\Delta ES})$ are updated every MCMC iteration. Notice that we have not placed priors on the spline regression coefficients $\boldsymbol{\beta_{ee}}$ and $\boldsymbol{\beta_{es}}$, or the linear regression coefficients $\boldsymbol{\gamma_{ee}}$ and $\boldsymbol{\gamma_{es}}$; this is because we will not be updating them in a fully Bayesian fashion. We discuss this more in the RJMCMC section.

2.4 Simulated Data

In this section we describe how we simulate data to mimic "real" observations, in order to perform a simulation study. Our simulated data need to be sufficiently complex and incorporate dependence in order to faithfully represent the distributions of true EE and EI, as well as gold standard measurements and inexpensive measurements. We try to simulate data that look similar to the EBS data. We need to simulate data for all the components in (2.3) as well as the latent variables $(X_i^{EE}, X_i^{\Delta ES})$. We explore estimation with measurement errors for the gold standard and inexpensive measurements under three different scenarios: normal errors, skewed errors, bimodal errors.

2.4.1 Demographic covariates

For this simulation, we used three covariates: gender, age, BMI. Using a total sample size of 300, we sampled 300 Bernoulli(0.5) to determine gender. Age was simulated from Uniform(20,40) to mimic the EBS. The BMI for an individual was simulated from a Normal(27,5). Let Z be the matrix of dimension 300×3 that links covariates to individuals.

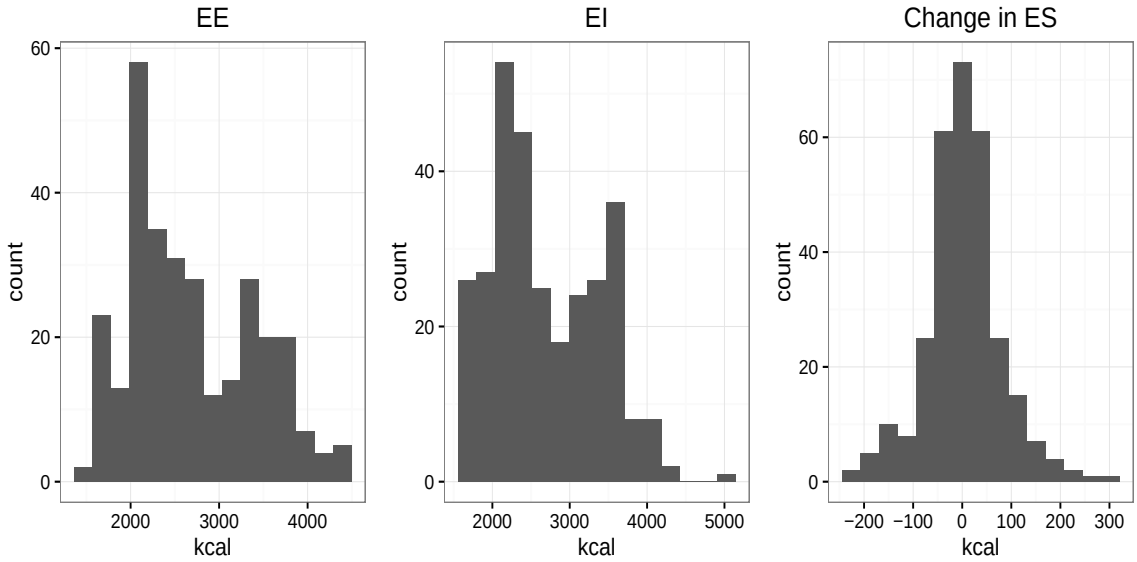

Figure 2.6: Distribution of simulated latent variables X from one simulated data set

2.4.2 Latent variables

We simulate (X_i^{EE}, X_i^{EI}) from a mixture of 5 bivariate t-distributions. Sixty observation pairs are simulated from five different bivariate t-distributions. Simulation was carried out using the `rmvt` function in the `mvtnorm` package in `R`. The mean and standard deviation of the two-dimensional vector for each of the five t-distributions is given in Table 2.2. The scale matrix for each of the five t-distributions is constant for standard deviations and correlation and are chosen empirically. The degrees of freedom is equal to five.

We let the correlation between EE and EI be 0.4376 as calculated from the EBS data. The values used for the vector $\gamma_{ee} = (300, 14, -7)$ and $\gamma_{es} = (-200, 8, -5)$ for gender, BMI and age, respectively. The matrix $Z_{comp,k}$ is a 60x3 matrix of covariate values corresponding to the 60 individuals in component k. We compute $X_i^{\Delta ES}$ using the energy balance equation in (2.1). Figure 2.6 shows histograms for the latent variables for one simulated data set. Figure 2.7 shows contours of the bivariate distributions.

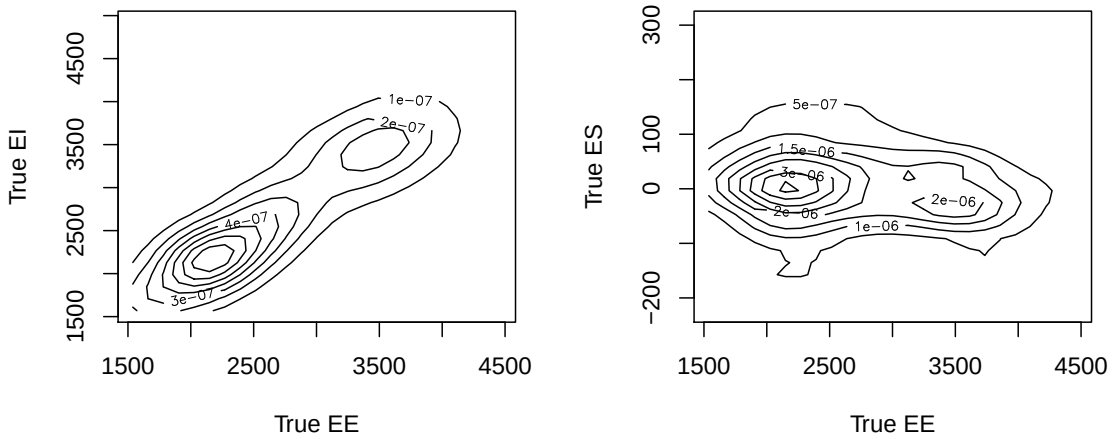

Figure 2.7: Contour plots of bivariate distributions of one simulated data set for latent variables. X^{EE} vs X^{EI} on left and X^{EE} vs $X^{\Delta ES}$ on the right panel

Table 2.2: Mean and Variances for X

	Mean EE	Mean EI	Std Dev EE	Std Dev EI	Cor(EE,EI)
Component 1	$1500 + \gamma_{ee} Z_{comp,1}$	$1630 + \gamma_{ei} Z_{comp,1}$	50	70	0.4376
Component 2	$1900 + \gamma_{ee} Z_{comp,2}$	$2030 + \gamma_{ei} Z_{comp,2}$	80	95	0.4376
Component 3	$2100 + \gamma_{ee} Z_{comp,3}$	$2180 + \gamma_{ei} Z_{comp,3}$	60	90	0.4376
Component 4	$2900 + \gamma_{ee} Z_{comp,4}$	$3030 + \gamma_{ei} Z_{comp,4}$	400	410	0.4376
Component 5	$3200 + \gamma_{ee} Z_{comp,5}$	$3330 + \gamma_{ei} Z_{comp,5}$	220	250	0.4376

2.4.3 Within person variability

For the gold standard measurements, let

$$\begin{aligned} \nu_{ij}^{EE} &= u_{ij}^{EE} + \delta_{ij}^{EE} \\ \nu_{ij}^{\Delta ES} &= u_{ij}^{\Delta ES} + \delta_{ij}^{\Delta ES}, \end{aligned} \tag{2.60}$$

where u^{EE} represents the measurement error in DLW and $u^{\Delta ES}$ represents the measurement error in DXA. Above, δ_{ij}^{EE} represents the within person deviation in EE for person i during time period j from the person's true mean, and $\delta_{ij}^{\Delta ES}$ similarly represents the within person deviation in ΔES for person i during time period j from the person's true mean. For the inexpensive measurements there is a slightly different setup. The within person variability gets added to each individuals' *usual* values of EE and ΔES and thus is affected by the functions $m_{\cdot}(\cdot)$. Therefore we add these within person variation terms δ to the *usual* X values we simulated to get:

$$\begin{aligned} X_{ij}^{EE} &= X_{i}^{EE} + \delta_{ij}^{EE} \\ X_{ij}^{\Delta ES} &= X_{i}^{\Delta ES} + \delta_{ij}^{\Delta ES}, \end{aligned} \tag{2.61}$$

and the functions $m_{\cdot}(\cdot)$ depend on $X_{ij}^{\cdot}$.

The pairs $(\delta_{ij}^{EE}, \delta_{ij}^{\Delta ES})$ are simulated jointly but independently across time and individual. We simulate the within person variability terms $(\delta_{ij}^{EE}, \delta_{ij}^{\Delta ES})$ according to the bivariate distribution below:

$$(\delta_{ij}^{EE}, \delta_{ij}^{\Delta ES}) \overset{iid}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\delta^{EE}} & \rho_\delta \sigma_{\delta^{EE}} \sigma_{\delta^{\Delta ES}} \\ \rho_\delta \sigma_{\delta^{EE}} \sigma_{\delta^{\Delta ES}} & \sigma^2_{\delta^{\Delta ES}} \end{bmatrix} \right), \tag{2.62}$$

with $\sigma_{\delta^{EE}} = 150$, $\sigma_{\delta^{\Delta ES}} = 50$, and $\rho_\delta = -0.266$.

2.4.4 Gold standard measurements

We assume that DLW and DXA are unbiased measurements of EE and ΔES, respectively. These measurements are simulated according to (2.6),(2.7) where we further brake down ν as in

(2.60). The u term represents the measurement error components we still need to specify and δ represents the within person component of the error which we have already discussed. We assume that the u terms are independent within and across individuals as well as of all δ and X.

From these simulated values, we then get simulated gold standard data $W_{ij}^{EE}, W_{ij}^{\Delta ES}$ via (2.6),(2.7). We generate measurement errors for the gold standard measurements (and for the inexpensive measurements) from three different distributions: normal, skewed normal, and a bimodal mixture of two normals that is centered around 0. The specification for the normal errors are given in (2.63). The skewed normal distribution was introduced by Fernandez and Steel (1998). The measurement error for this distribution is specified by (2.64). We also explore a bimodal mixture of two normal distributions as a generative model for measurement errors. We constructed it such that it is symmetric and centered around 0. This distribution is specified by (2.65). Parameters were chosen such that the means of all error distributions are 0, and the variances for each distribution is the same within EE errors and within ΔES errors. We simulate from the skewed normal distribution using the R package `fGarch`. Figure (2.8) shows density plots of all three distributions using our specified standard deviation and skew parameters. The parameters from which we generated measurement errors are shown below:

$$\begin{aligned} u_{ij}^{EE} &\overset{iid}{\sim} N(0, \sigma^2_{\nu^{EE}}) \\ u_{ij}^{\Delta ES} &\overset{iid}{\sim} N(0, \sigma^2_{\nu^{\Delta ES}}) \end{aligned} \tag{2.63}$$

$$\begin{aligned} u_{ij}^{EE} &\overset{iid}{\sim} SN(0, \sigma^2_{\nu^{EE}}, 10) \\ u_{ij}^{\Delta ES} &\overset{iid}{\sim} SN(0, \sigma^2_{\nu^{\Delta ES}}, 10) \end{aligned} \tag{2.64}$$

$$\begin{aligned} u_{ij}^{EE} &\overset{iid}{\sim} 0.5 \times N(-175, \sigma^2_{\nu^{EE}} - 175^2) + 0.5 \times N(175, \sigma^2_{\nu^{EE}} - 175^2) \equiv BM(175, \sigma^2_{\nu^{EE}}) \\ u_{ij}^{\Delta ES} &\overset{iid}{\sim} 0.5 \times N(-45, \sigma^2_{\nu^{\Delta ES}} - 45^2) + 0.5 \times N(45, \sigma^2_{\nu^{\Delta ES}} - 45^2) \equiv BM(45, \sigma^2_{\nu^{\Delta ES}}), \end{aligned} \tag{2.65}$$

$$\sigma_{\nu^{EE}} = 200, \sigma_{\nu^{\Delta ES}} = 53 \tag{2.66}$$

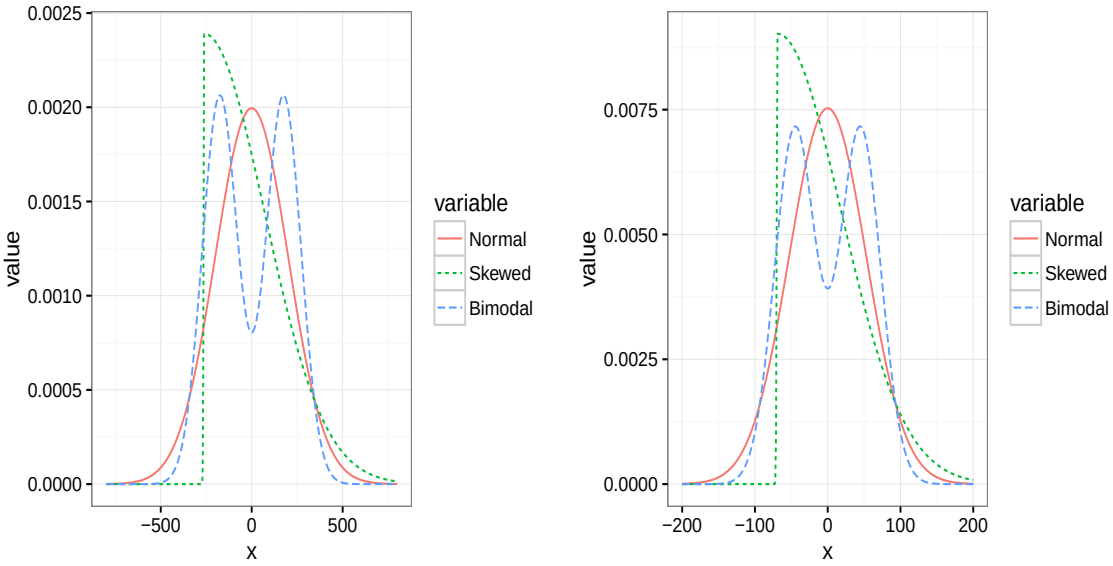

Figure 2.8: Measurement Error Distributions, EE on left and ΔES on right

2.4.5 Inexpensive measurements

We simulate observations from the inexpensive measurements according to (2.4), (2.5). Equation (2.61) shows how we break down the error term from (2.4),(2.5). We use the same simulated values of δ that we used for the gold standard measurements.

We generate the measurement error terms e in a similar fashion as in the last section. We assume that the errors are independent within and across subjects as well as mutually independent with all δ, X and Z terms. We draw these errors from densities that are similar to those in the previous section:

$$\begin{aligned} e_{ij}^{EE} &\overset{iid}{\sim} N(0, \sigma^2_{\epsilon^{EE}}) \\ e_{ij}^{\Delta ES} &\overset{iid}{\sim} N(0, \sigma^2_{\epsilon^{\Delta ES}}), \end{aligned} \tag{2.67}$$

$$\begin{aligned} e_{ij}^{EE} &\overset{iid}{\sim} SN(0, \sigma^2_{\epsilon^{EE}}, 10) \\ e_{ij}^{\Delta ES} &\overset{iid}{\sim} SN(0, \sigma^2_{\epsilon^{\Delta ES}}, 10), \end{aligned} \tag{2.68}$$

$$\begin{aligned} e_{ij}^{EE} &\overset{iid}{\sim} 0.5 \times N(-350, \sigma^2_{\epsilon^{EE}} - 350^2) + 0.5 \times N(350, \sigma^2_{\epsilon^{EE}} - 350^2) \equiv BM(350, \sigma^2_{\epsilon^{EE}}) \\ e_{ij}^{\Delta ES} &\overset{iid}{\sim} 0.5 \times N(-190, \sigma^2_{\epsilon^{\Delta ES}} - 190^2) + 0.5 \times N(190, \sigma^2_{\epsilon^{\Delta ES}} - 190^2) \equiv BM(190, \sigma^2_{\epsilon^{\Delta ES}}), \end{aligned} \tag{2.69}$$

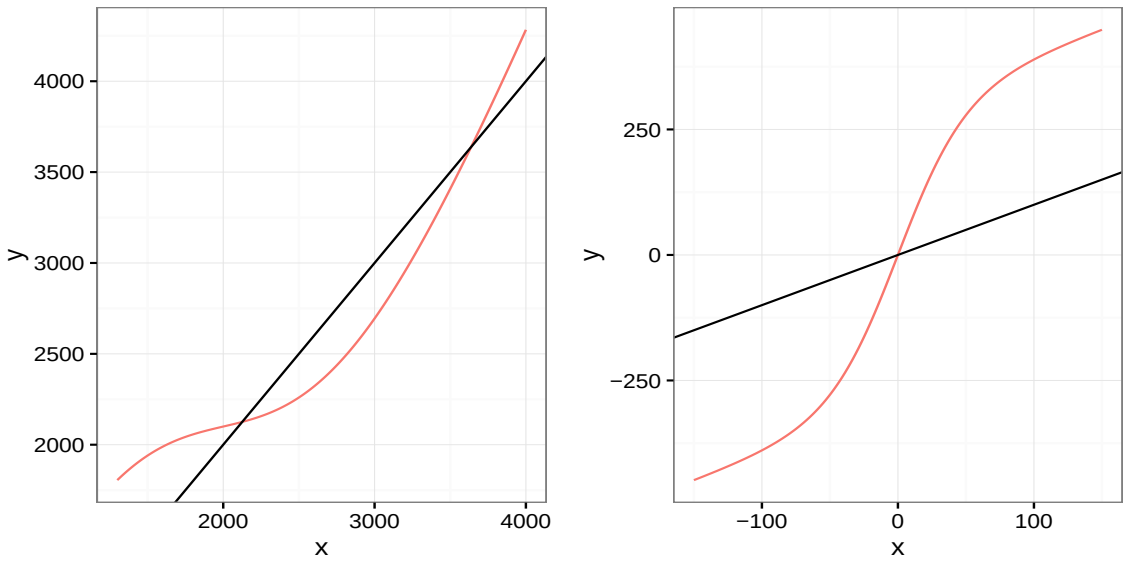

Figure 2.9: Plot of nonlinear functions $m_{ee}()$ (left) and $m_{es}()$ (right), and Y=X is black for reference to unbiased measurement

$$\sigma_{\epsilon^{EE}} = 380, \sigma_{\epsilon^{\Delta ES}} = 210. \tag{2.70}$$

In contrast to the gold standard measurements which we assume are unbiased, we now add bias to the inexpensive measurements. The bias is introduced via the functions m_{ee} and m_{es}. For these simulated data, we let:

$$m_{ee}(X, Z) = 2X_i^{EE} - \frac{4000}{1 + e^{-0.002X_i^{EE} - 2200}}, \tag{2.71}$$

$$m_{es}(X, Z) = \frac{1000}{1 + e^{-0.04X_i^{\Delta ES}}} - 2000 + X_i^{\Delta ES}. \tag{2.72}$$

We then add $Z_i\boldsymbol{\gamma}$. to the simulated inexpensive measures of EE and ΔES. Figure 2.9 shows $m_{ee}(\cdot)$ on the left and $m_{es}(\cdot)$ on the right both against a $y = x$ line for comparison.

2.5 Estimation

We adopt a Bayesian approach to estimation in this problem, and therefore, our goal is to estimate the joint posterior distribution of all parameters and latent variables in the model. In

our case, the joint posterior distribution is $p(\boldsymbol{\theta}, X^{EE}, X^{\Delta ES}|W^{EE}, W^{\Delta ES}, Y^{EE}, Y^{\Delta ES}, Z)$. We use Markov Chain Monte Carlo (MCMC) methods to approximate the posterior distribution. For the naïve and LMEM models, we used JAGS to simulate draws from the posterior distribution. This was simple to implement and was relatively quick to sample. In order to fit free knot splines which allow for dimension change, we must use Reversible Jump MCMC which requires a more complex sampler. We first present the Gibbs sampler with the reversible jump as a single step, and in the next subsection we describe the reversible jump step in more detail.

2.5.1 Gibbs algorithm for the spline model

Here we present our Gibbs algorithm to sample from the joint posterior distribution.

1. Set starting values. Let $i = 1, ..., n$ index individuals, and $j = 1, ..., J$ index replicates within individual. Assume for simplicity each individual has the same number of replicates.

2. Run the MCMC below B times. Every b iterations, update tuning parameters C_{X_i}, the covariance for the random walk proposal distribution by setting

$$C_{X_i} = \frac{2.4^2}{2} \frac{1}{B-1} \sum_{b=1}^{B} (X_i^{EE(b)} - \bar{X}^{EE})(X_i^{\Delta ES(b)} - \bar{X}^{\Delta ES}). \tag{2.73}$$

   Repeat desired number of times, then discard burn-in draws and continue using new tuning values. (See Gelman et al. (2014), pg. 290)

3. For iteration k=1,...,K, sample from the full conditional distributions:

   (a) $\{\zeta_i^{(k)} : i = 1, ..., n\} \mid \cdot$ where $\mathrm{P}(\zeta_i^{(k)} = h) = \frac{\pi_h^{(k-1)} N(\mu_h^{(k-1)}, \Sigma_h^{(k-1)})}{\sum_{h=1}^{H} \pi_h^{(k-1)} N(\mu_h^{(k-1)}, \Sigma_h^{(k-1)})}$.

   (b) Draw $\{V_h^{(k)} : h = 1, ..., H-1\}|\cdot \overset{ind}{\sim} Beta(1 + n_h, \alpha + n_h')$,
   With $V_H^{(k)} = 1$
   $n_h = \sum_{i=1}^{n} I(\zeta_i^{(k)} = h)$

$n'_h = \sum_{h'=h+1}^{H} n_{h'}$.

(c) Calculate $\{\pi_i^{(k)} : i = 1, ..., n\} = V_i^{(k)} \prod_{\ell<i}(1 - V_\ell^{(k)})$.

(d) Draw $\{\Sigma_h^{(k)}, h = 1, ..., H\} \mid \cdot \overset{ind}{\sim} Inv{-}Wish(d+n_h, \psi+(\mathbf{X_{i,h}^{(k-1)}} - \mu_\mathbf{h}^{(\mathbf{k-1})})'(\mathbf{X_{i,h}^{(k-1)}} - \mu_\mathbf{h}^{(\mathbf{k-1})}))$,

with

$\mathbf{X_{i,h}} \equiv \mathbf{X_i} I(\zeta_i^{(k)} = h)$

$\mathbf{X_i} = (X_i^{EE}, X_i^{\Delta ES})$

$\mu_\mathbf{h} = (\mu_{EE,h}, \mu_{\Delta ES,h})$.

(e) Draw $\{(\mu_{EE,h}^{(k)}, \mu_{\Delta ES,h}^{(k)}), h = 1, ..., H\ \} \mid \cdot \overset{ind}{\sim} N(M'_\mu, C'_\mu)$,

with

$C'_\mu = (C_\mu^{-1} + n_h \Sigma_h^{-1(k)})^{-1}$

$M'_\mu = C'_\mu (C_\mu^{-1} M + n_h \Sigma_h^{-1(k)} \bar{\mathbf{X}}^{(\mathbf{k-1})})$

$\bar{\mathbf{X}} = \frac{1}{n_h} \sum_{i=1}^{n} \mathbf{X_{i,h}}$.

(f) Update $\mathbf{X_i}$ with a random walk

$\{\mathbf{X_i^{(k)}} : i = 1, ..., n\} \mid \cdot$ for $i = 1, ...n$ sample $\mathbf{X_i^*}$ from $N(\mathbf{X_i^{(k)}}, C_{X_i})$,

and set $\mathbf{X_i^{(k)}} = \mathbf{X_i^*}$ with probability α_{X_i}, otherwise set $\mathbf{X_i^{(k)}} = \mathbf{X_i^{(k-1)}}$,

where

$\alpha_{X_i} = min\left(1, \frac{f(\mathbf{X_i^*}|\cdot)}{f(\mathbf{X_i^{(k-1)}}|\cdot)}\right)$.

(g) Update $\boldsymbol{\beta_{ee}}, \boldsymbol{\beta_{es}}, k_{ee}, k_{es}, r_{ee}, r_{es}, \boldsymbol{\gamma_{ee}}, \boldsymbol{\gamma_{es}}$ using RJMCMC described in next section. Calculate $s_{ee}(\mathbf{X^{EE(k)}}; \boldsymbol{\beta_{ee}^{(k)}})$ and $s_{es}(\mathbf{X^{\Delta ES(k)}}; \boldsymbol{\beta_{es}^{(k)}})$.

(h) Draw $\sigma^{2(k)}_{\epsilon^{EE}}|\cdot \sim$
$IG\left(a_{yee} + J \times \frac{n}{2}, b_{yee} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}\left(Y_{ij}^{EE} - s_{ee}(X_i^{EE(k)}; \boldsymbol{\beta_{ee}^{(k)}}) - \boldsymbol{\gamma_{ee}}^{(k)'} Z_i\right)^2\right).$

(i) Draw $\sigma^{2(k)}_{\epsilon^{\Delta ES}}|\cdot \sim$
$IG\left(a_{yes} + J \times \frac{n}{2}, b_{yes} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}\left(Y_{ij}^{\Delta ES} - s_{es}(X_i^{\Delta ES(k)}; \boldsymbol{\beta_{es}^{(k)}}) - \boldsymbol{\gamma_{es}}^{(k)'} Z_i\right)^2\right).$

(j) Draw $\sigma^{2(k)}_{\nu^{EE}}|\cdot \sim$
$IG\left(a_{wee} + J \times \frac{n}{2}, b_{wee} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}\left(W_{ij}^{EE} - X_i^{EE(k)}\right)^2\right).$

(k) Draw $\sigma^{2(k)}_{\nu^{\Delta ES}}|\cdot \sim$
$IG\left(a_{wes} + J \times \frac{n}{2}, b_{wes} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{J}\left(W_{ij}^{\Delta ES} - X_i^{\Delta ES(k)}\right)^2\right).$

The steps above describe one iteration in the algorithm.

2.5.2 Reversible jump MCMC

Green (1995) introduced RJMCMC as a means to jump between parameter spaces that have different dimensions within an MCMC algorithm. There have been a number of different approaches to Bayesian estimation of free knot splines using RJMCMC (DiMatteo et al. (2001), Lindstrom (2002), Johnson (2007), Denison et al. (1998)). For the most part, we adopt the approach of Denison et al. This approach performs well relative to the fully Bayesian approach of DiMatteo et al. for smooth and not highly complex functions when an appropriate adjustment is made (that was pointed out in DiMatteo et al.). We do not expect the mean function in our model to be highly complex, so the approach of Denison et al. is likely to perform well in our problem.

Consider the situation where we wish to estimate a regression spline in the following setup

$$\begin{aligned} y_i &= f(x_i; \boldsymbol{\beta}) + \epsilon_i, \\ \epsilon_i &\overset{iid}{\sim} N(0, \sigma^2). \end{aligned} \tag{2.74}$$

where the $y_i, i = 1, ...n$ are the response variable and $x_i, i = 1, ...n$ is the observed covariate. We do not want to specify the number or location of the knots, but rather estimate them from the data. Under the specifications of Denison et al., in RJMCMC for free knot splines there are three possible transitions:

1. Birth of a knot
2. Death of a knot
3. Movement of a knot,

where in either of the first two transitions the dimension of the parameter space changes. The transition that is proposed depends upon the prior for the number of knots k. Using the notation of Denison et al., let the prior probability for k knots be $p(k)$. Then the probability of attempting a birth step, death step or move step are b_k, d_k, η_k, respectively, where $\eta_k = 1 - b_k - d_k$, and

$$\begin{aligned} b_k &= c \times min\left(1, \frac{p(k+1)}{p(k)}\right) \\ d_k &= c \times min\left(1, \frac{p(k)}{p(k+1)}\right), \end{aligned} \tag{2.75}$$

where $0 \leq c \leq 1/2$ to ensure $b_k + d_k \leq 1$. If the birth step transition is chosen, a new knot is proposed from the set of available new knots. Denison et al. puts a discrete uniform prior on knot locations so that only observed locations can be knot locations. A proposed new knot is chosen at random from the set $\mathcal{A}$=$\{x_i, i = 1, ...n : x_i$ is not currently a knot or within $\ell + 1$ locations of a current knot$\}$. Here, ℓ is the order of the splines, so in our case ℓ =3. If the death step transition is chosen, an existing knot is picked at random and removed. If the movement transition step is chosen, a current knot is picked at random and moved to a new location at random from the set $\mathcal{A}$. Once new knots are chosen, we construct a spline basis matrix using the observed locations $\{x_i, i = 1, ...n\}$ and the positions of the knots. We then use OLS to estimate the spline parameters $\boldsymbol{\beta}$. Although we could place priors on the spline regression parameters, this adds significant computational burden and results have been similar when comparing estimation

with OLS versus fully Bayesian estimation as long as functions are smooth and not highly complex (Denison et al. (1998), DiMatteo et al. (2001)). The acceptance probability for each step is of the form given by Green (1995):

$$\alpha = min(1, \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio}). \tag{2.76}$$

Denison et al. give simplified acceptance probabilities for each of the three transitions under a Poisson prior for number of knots and a discrete uniform prior for knot locations. In the regression setup in Denison et al., they then sample variance components using a Gibbs step.

2.5.3 RJMCMC implementation

The model that was fitted in Denison et al. (1998) is simpler than our model (2.74) in three respects: (i) our regression model is part of a larger hierarchical model, (ii) we have an additional linear component in the mean function, (iii) the covariate value x_i is a latent variable . The first issue does not present much of a problem thanks to conditional independence assumptions. For issue (ii) we propose to update the linear coefficient parameters at the same time as we update the spline coefficient parameters, using OLS. We understand that this is not a fully Bayesian approach, but we anticipate results to be similar with much less computational burden.

Issue (iii) is not as simple. In the free knot spline setup in Denison et al., they have a regression model of the form (2.74) where y is the observed response and x is the observed covariate. In our case we do not observe x; rather, it is a latent variable that we draw from its full conditional distribution using the Gibbs algorithm described in Section 2.5.1. Because we sample the x's via the Gibbs algorithm, the basis matrix has to be adjusted in every iteration. We calculate the basis matrix using the current value of the latent variables and the proposed knots. If we accept the proposal, then the knots, spline coefficients, and consequently the estimate of $s_\cdot(X^{\cdot(k)}, \boldsymbol{\beta}_\cdot^{(k)})$ are all updated. The challenge is what to do when we reject the candidate draw. It seems reasonable to keep the current knot locations and the current spline regression parameters, but keeping the current predictions of $s_\cdot(X^{\cdot(k)}, \boldsymbol{\beta}_\cdot^{(k)})$ does not make sense because they are based on $X^{\cdot(k)}$, which

is updated in every iteration. We propose the following protocol: if we reject the RJ candidate, we keep the current knots and spline coefficients, calculate a new basis matrix of X based on the current values of $X^{\cdot(k)}$ and the current knots, and then compute the predicted values by multiplying the basis matrix by the current spline coefficients. An additional complication is that X has to do with the birth step and the set of available knots $\mathcal{A}$ from which to choose . We want to pick randomly from the set of X values which we drew in the current iteration of the MCMC, but we do not want a knot to be bigger than or smaller than the maximum or minimum value of X. To avoid this problem, we exclude the three smallest and the three largest values of X from $\mathcal{A}$. We do not expect this to be a major constraint because we are less concerned with flexibility of the function at its extremes and therefore do not want to allow lots of knots at the extremes.

This algorithm is run independently for EE and ΔES mean regression functions because of the conditional independence assumptions. Because this algorithm is the same for EE and ΔES, we omit subscripts and superscripts below. The reversible jump step that goes into our overall Gibbs algorithm is then:

1. Calculate b_k and d_k according to (2.75).

2. Select birth, death, or move step with probabilities $b_k, d_k, 1 - b_k - d_k$ respectively.

3. **Knot Changes**

   If birth step:

   (a) Select a new knot location at random from the set $\mathcal{A}$ and join with current knots $r^{(k-1)}$ to create the proposed knot locations r^*.

   If death step:

(b) Sample one knot location from $r^{(k-1)}$ at random and remove it to create the proposed knot locations r^*.

If move step:

(c) Sample one knot location from $r^{(k-1)}$ at random, and change it to a new knot location at random from the set $\mathcal{A}$ to create the proposed knot locations r^*.

4. Calculate the spline basis matrix $B^*(\mathbf{X^{(k)}})$ using $\mathbf{X^{(k)}}$ and proposed knot locations r^*.

5. Calculate proposed spline and linear regression coefficients $\boldsymbol{\beta^*}, \boldsymbol{\gamma^*}$ by using OLS to regress $\bar{\mathbf{Y}}$ on $B^*(\mathbf{X}^{\cdot(\mathbf{k})}) + \mathbf{Z}$
$\bar{\mathbf{Y}} = \left\{ \frac{1}{J} \sum_{j=1}^{J} Y_{ij} : i = 1, ..., n \right\}$.

6. Accept proposed knots and coefficients with probability α, with α shown below:

$$\alpha_{birth} = min\left(1, \text{Likelihood ratio} \times \frac{n - Z(k)}{n}\right),$$
$$\alpha_{death} = min\left(1, \text{Likelihood ratio} \times \frac{n}{n - Z(k)}\right),$$
$$\alpha_{move} = min\left(1, \text{Likelihood ratio}\right),$$

Otherwise set $r^{(k)} = r^{(k-1)}$, $\boldsymbol{\beta^{(k)}} = \boldsymbol{\beta^{(k-1)}}$, $\boldsymbol{\gamma^{(k)}} = \boldsymbol{\gamma^{(k-1)}}$, and

$$Z(k) = 2(\ell + 1) + k.(2\ell + 1),$$
$$k = length(r^{(k-1)}).$$

7. Compute mean function $s(\mathbf{X^{(k)}}; \boldsymbol{\beta^{(k)}})$ using spline basis matrix $B'(\mathbf{X^{(k)}})$ and $r^{(k)}$

2.6 Simulation Study

In this section we describe a simulation study that we carried out, to check the performance of the models we propose. We are interested in the predictive performance of the models because our

main goal is to develop a calibration tool. We are also interested in evaluating the robustness of the model to departures of the errors from the standard normality assumption, which is why we simulate errors from two alternative error distributions. We present performance measures such as predicted mean squared error (PMSE) for the regression function in question as well as posterior means and posterior standard deviations for parameters of interest.

2.6.1 Setup

We simulated 200 data sets each for normal, skewed, and bimodal errors for both 2 and 4 replicate measurements per individual. The number of individuals is 300 in all cases. Preliminary analysis suggest that the number of replicates per individual has a stronger impact on performance than the number of individuals.

Although we would like to be as flexible as possible with our distributional assumptions on the bivariate latent variables, we also want a model that produces estimates with low MSE given the data constraints of our application. In practice, it is difficult to obtain more than two replicate measurements on an individual, at least when using the gold standard measurements. To obtain three replicate measurements in the DLW case amounts to asking participants to be in the study during three two-week periods and undergo a washout time in between. During the simulation study, we found that the Dirichlet Process prior on the latent variables produced unstable results in parameter estimates and low acceptance rates of proposals in the random walk Metropolis-Hastings algorithm when we only had two replicate observations per person. Results were stable however, when four replicates per person were available. In real-world application it is unrealistic to obtain more than two replicates per person. Because of this issue, we fit a fourth model using a bivariate normal distribution for the prior of latent variables instead of the Dirichlet Process prior while still using splines for the regression functions. We refer to this model as SMEMN. The MCMC has a minor change in the Gibbs step (steps (a)-(c) are eliminated and step (d) no longer depends on grouping h).

We set the values of the hyperparameters as follows: $M_{\beta_{0,ee}} = M_{\beta_{0,es}} = 0, C_{\beta_{0,ee}} = C_{\beta_{0,es}} = 100000, M_{\beta_{1,ee}} = M_{\beta_{1,es}} = 1, C_{\beta_{1,ee}} = C_{\beta_{1,es}} = 100000, M_{\gamma_{ee}} = M_{\gamma_{es}} = 0, C_{\gamma_{ee}} = C_{\gamma_{es}} = 100000, a_{yee} = a_{yes} = a_{wee} = a_{wes} = b_{yee} = b_{yes} = b_{wee} = b_{wes} = 0.1, \psi = I_{2\times 2}, d = 3, M_{\mu} = (2400, 0), C_{\mu} = diag(100000, 100000), \lambda_{ee} = \lambda_{es} = 1$. We ran the MCMC for 3 chains of 12,000 iterations, using the first 2000 as burn in, and convergence for all models was fast as indicated by trace plots and Gelman-Rubin diagnostics less than 1.04.

2.6.2 Results

Tables 2.3,2.4,2.5 show results averaged over 200 Monte Carlo samples, for normal, skewed, and bimodal errors, respectively. The asterisk next to the truth for the measurement error with respect to the inexpensive measurements indicates that this is a Monte Carlo approximation to the truth. Recall that we included within person variation in the functions $m.(\cdot)$, but in our model we use the working assumption that the additive error term accounts for both within person variability and measurement error. Because we cannot directly extract the value from the function, we approximate it by generating 10,000 data sets and removing the mean function from the inexpensive observations, and then calculating the standard deviation of the residual. We then averaged those standard deviation estimates to get the one reported in the table.

Across all models and error types, the linear coefficients are estimated largely without bias. This is not too surprising since these covariates are measured without error. The biases and standard errors are slightly smaller for models SMEMN and SMEMDP, however. All three measurement error models perform about the same when assessing the measurement error in the gold standard instruments. When errors are generated from a bimodal distribution, estimated error variances are biased toward zero. This is true for the measurement error in the inexpensive measurements as well. The SMEMN and SMEMDP models produce similar results for the estimates of variance measurement error of inexpensive measurements. Estimates are good for EE and ΔES when errors are normal, but biased low for ΔES for both skewed and bimodal errors. Both the naïve model and the linear measurement error model result in estimated measurement error standard deviations

for the inexpensive measurement that are too large under normal errors and skewed errors for EE. When the departure from normality is significant (bimodal error distribution) unbiasedly estimating the measurement error variance can be challenging.

Figure 2.10 shows boxplots of the log mean PMSE for each simulation for each model under each type of error distribution for both EE and ΔES for 2 and 4 replicates. There is a consistent decreasing pattern from simpler to most complex in terms of the models. First, the naïve model does much worse than the same model which accounts for measurement error. The naïve model and the linear measurement error model perform much worse than the models with free knot splines in terms of PMSE. There is not a large difference between the SMEMN and SMEMDP model in terms of PMSE, but the SMEMDP model generally does better. The question is whether the small improvement is worth the increase in model complexity. We think that the answer is no for two reasons: (i) our main focus with this model is calibrating the inexpensive measurements and not necessarily conducting inference at the latent variable level, and (ii) the DP approach is reliable only situations when we have four replicates, which for gold standard measurements, is unrealistic in practice. Given that the latent variables were non-normal, we might explore whether a simple bivariate normal model might fit these latent data. We address this question in the context of an example in the next section. Because the main focus is to calibrate inexpensive measurements, the simulation results are promising.

2.6.3 A closer look at one simulated dataset

In this section we provide an example analysis using one of the simulated data sets. This gives an idea of the type of results that we might obtain for a "real" data sets when data become available. We provide posterior quantiles to understand uncertainty in all estimates. We also provide posterior predictive assessments for our modeling assumptions. The posterior predictive distributions of interest are given in (4.65),(2.78),(2.79).

$$p(Y^*|W,Y,Z) = \int\int p(Y^*|\theta,X,Z)p(\theta,X|Y,W)d\theta dX \tag{2.77}$$

$$p(W^*|W,Y,Z) = \int\int p(W^*|\theta,X,Z)p(\theta,X|Y,W)d\theta dX \tag{2.78}$$

$$p(X^*|W,Y,Z) = \int\int p(X^*|\theta)p(\theta,X|Y,W)d\theta dX. \tag{2.79}$$

Table 2.3: Summary of Simulation under Normal Errors for naïve, LMEM, SMEMN, SMEMDP Models, respectively

Naïve	σ_{yee}		σ_{yes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	477.65	473.42	347.49	354.85	254.67	248.66	14.88	14.03	-4.14	-5.29	-200.53	-199.39	7.91	8.31	-4.94	-5.13
Std Err	17.82	19.24	9.43	6.80	43.65	36.10	4.33	3.62	3.37	3.06	28.19	22.83	2.90	2.13	2.35	1.73
Bias	72.15	67.92	13.49	20.85	-45.33	-51.34	0.88	0.03	2.86	1.71	-0.53	0.61	-0.09	0.31	0.06	-0.13
Truth	405.50	405.50	334.00	334.00	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

LMEM	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	444.34	446.63	320.41	338.26	255.70	255.85	69.18	71.81	249.50	240.63	14.30	13.67	-4.50	-5.28	-199.27	-198.49	7.91	8.29	-4.91	-5.15
Std Err	16.84	14.22	10.76	7.53	10.74	6.33	2.27	1.56	43.44	37.02	4.25	3.60	3.38	3.04	28.31	22.96	2.91	2.12	2.30	1.71
Bias	38.84	41.13	-13.59	4.26	5.70	5.85	-3.68	-1.05	-50.50	-59.37	0.30	-0.33	2.50	1.72	0.73	1.51	-0.09	0.29	0.09	-0.15
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

SMEMN	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	393.69	400.55	313.47	331.79	246.81	248.93	67.61	71.04	293.16	294.61	14.17	14.11	-6.86	-6.78	-200.07	-200.86	8.00	8.04	-4.95	-5.26
Std Err	11.52	8.29	12.00	8.27	8.78	5.79	2.32	1.66	36.16	26.19	3.50	2.42	2.86	2.16	26.86	19.44	2.49	2.06	2.11	1.45
Bias	-11.81	-4.95	-20.53	-2.21	-3.19	-1.07	-5.25	-1.82	-6.84	-5.39	0.17	0.11	0.14	0.22	-0.07	-0.86	0.00	0.04	0.05	-0.26
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

SMEMDP	σ_{yee}	σ_{yes}	σ_{wee}	σ_{wes}	$\gamma_{1,ee}$	$\gamma_{2,ee}$	$\gamma_{3,ee}$	$\gamma_{1,es}$	$\gamma_{2,es}$	$\gamma_{3,es}$
Replicates	4	4	4	4	4	4	4	4	4	4
Mean Est	400.13	331.38	248.94	70.85	297.62	14.16	-7.11	-198.04	8.12	-4.94
Std Err	8.25	8.43	6.05	1.62	26.53	2.72	2.08	18.11	1.85	1.64
Bias	-5.37	-2.62	-1.06	-2.01	-2.38	0.16	-0.11	1.96	0.12	0.06
Truth	405.50	334.00	250.00	72.86	300.00	14.00	-7.00	-200.00	8.00	-5.00

Table 2.4: Summary of Simulation under Skewed Errors for naïve, LMEM, SMEMN, SMEMDP Models, respectively

Naïve	σ_{yee}		σ_{yes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	473.87	466.80	311.66	317.05	255.46	250.72	14.08	13.23	-5.05	-6.04	-197.98	-200.46	7.79	7.91	-4.98	-4.92
Std Err	17.36	13.06	9.08	6.64	40.99	32.70	3.99	3.53	3.38	3.02	24.90	19.73	2.45	1.78	2.28	1.67
Bias	68.37	61.30	-22.34	-16.95	-44.54	-49.28	0.08	-0.77	1.95	0.96	2.02	-0.46	-0.21	-0.09	0.02	0.08
Truth	405.50	405.50	334.00	334.00	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

LMEM	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	432.67	435.88	289.34	304.25	254.13	255.00	69.67	71.68	251.72	244.04	13.38	12.88	-5.46	-6.10	-196.76	-200.05	7.82	7.89	-4.96	-4.94
Std Err	16.70	12.10	9.90	7.05	11.99	6.27	2.63	1.80	41.34	33.01	3.95	3.50	3.35	3.00	25.32	19.53	2.39	1.77	2.25	1.67
Bias	27.17	30.38	-44.66	-29.75	4.13	5.00	-3.20	-1.18	-48.28	-55.96	-0.62	-1.12	1.54	0.90	3.24	-0.05	-0.18	-0.11	0.04	0.06
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

SMEMN	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	393.56	402.91	282.68	298.37	247.00	248.83	68.93	71.23	306.58	305.65	14.27	14.44	-6.95	-7.14	-197.44	-198.85	7.76	8.04	-4.92	-4.86
Std Err	13.33	9.56	10.96	7.47	9.54	6.33	2.56	1.69	36.09	27.06	3.35	2.57	2.77	2.31	24.41	17.81	2.42	1.73	2.10	1.47
Bias	-11.94	-2.59	-51.32	-35.63	-3.00	-1.17	-3.93	-1.63	6.58	5.65	0.27	0.44	0.05	-0.14	2.56	1.15	-0.24	0.04	0.08	0.14
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

SMEMDP	σ_{yee}	σ_{yes}	σ_{wee}	σ_{wes}	$\gamma_{1,ee}$	$\gamma_{2,ee}$	$\gamma_{3,ee}$	$\gamma_{1,es}$	$\gamma_{2,es}$	$\gamma_{3,es}$
Replicates	4	4	4	4	4	4	4	4	4	4
Mean Est	403.10	298.16	248.70	71.11	313.22	14.13	-7.26	-199.31	8.10	-4.85
Std Err	8.25	7.01	6.67	1.57	27.80	2.37	2.13	17.70	1.54	1.54
Bias	-2.40	-35.84	-1.30	-1.75	13.22	0.13	-0.26	0.69	0.10	0.15
Truth	405.50	334.00	250.00	72.86	300.00	14.00	-7.00	-200.00	8.00	-5.00

Table 2.5: Summary of Simulation under Bimodal Errors for naïve, LMEM, SMEMN, SMEMDP Models, respectively

Naïve	σ_{yee}		σ_{yes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	342.63	344.52	233.52	246.69	227.08	221.48	12.15	12.22	-5.61	-5.52	-198.19	-199.20	8.01	8.13	-4.84	-4.99
Std Err	12.61	15.26	6.43	5.24	32.41	29.92	3.28	3.21	3.05	2.94	17.40	15.19	1.64	1.59	1.49	1.23
Bias	-62.87	-60.98	-100.48	-87.31	-72.92	-78.52	-1.85	-1.78	1.39	1.48	1.81	0.80	0.01	0.13	0.16	0.01
Truth	405.50	405.50	334.00	334.00	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

LMEM	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	264.16	265.06	220.65	237.60	207.41	207.86	52.92	56.11	216.24	203.84	11.50	11.52	-5.75	-5.38	-201.48	-201.44	7.90	8.07	-4.74	-4.92
Std Err	15.07	8.86	7.31	5.56	11.92	8.43	1.85	1.26	32.90	30.36	3.20	3.25	3.07	2.91	17.18	15.06	1.66	1.57	1.49	1.23
Bias	-141.34	-140.44	-113.35	-96.40	-42.59	-42.14	-19.94	-16.75	-83.76	-96.16	-2.50	-2.48	1.25	1.62	-1.48	-1.44	-0.10	0.07	0.26	0.08
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

SMEMN	σ_{yee}		σ_{yes}		σ_{wee}		σ_{wes}		$\gamma_{1,ee}$		$\gamma_{2,ee}$		$\gamma_{3,ee}$		$\gamma_{1,es}$		$\gamma_{2,es}$		$\gamma_{3,es}$	
Replicates	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4
Mean Est	257.20	256.24	217.62	235.09	182.30	189.38	51.86	55.39	220.14	211.17	12.25	11.96	-5.90	-5.81	-201.34	-200.20	7.87	7.98	-5.08	-4.95
Std Err	11.41	10.47	7.51	5.46	8.22	5.74	1.87	1.19	30.56	32.63	3.24	2.84	2.86	2.47	17.33	13.18	1.60	1.28	1.46	1.13
Bias	-148.30	-149.26	-116.38	-98.91	-67.70	-60.62	-21.01	-17.47	-79.86	-88.83	-1.75	-2.04	1.10	1.19	-1.34	-0.20	-0.13	-0.02	-0.08	0.05
Truth	405.50	405.50	334.00	334.00	250.00	250.00	72.86	72.86	300.00	300.00	14.00	14.00	-7.00	-7.00	-200.00	-200.00	8.00	8.00	-5.00	-5.00

SMEMDP	σ_{yee}	σ_{yes}	σ_{wee}	σ_{wes}	$\gamma_{1,ee}$	$\gamma_{2,ee}$	$\gamma_{3,ee}$	$\gamma_{1,es}$	$\gamma_{2,es}$	$\gamma_{3,es}$
Replicates	4	4	4	4	4	4	4	4	4	4
Mean Est	251.97	235.27	191.80	55.25	218.08	12.43	-5.67	-202.35	8.01	-5.14
Std Err	12.67	5.58	6.59	1.30	37.21	2.84	2.34	12.93	1.27	1.13
Bias	-153.53	-98.73	-58.20	-17.62	-81.92	-1.57	1.33	-2.35	0.01	-0.14
Truth	405.50	334.00	250.00	72.86	300.00	14.00	-7.00	-200.00	8.00	-5.00

To check whether our model can capture the true quantiles (min, 0.05, 0.1, 0.25, 0.75, 0.9, 0.95, max) of the bivariate latent vector distribution using both the inexpensive and gold standard measurements, we carry our a posterior predictive test. To do so, we behave like frequentists and generate replicate data sets from the model using the MCMC parameter draws. Doing this for many replicated datasets, or equivalently, many MCMC draws, we get distributions for the quantiles of interest. We then compare these distributions to the estimated quantile value, if the quantile estimated from the "observed" dataset lies well within the distribution of estimated quantiles obtained from replicate datasets, this suggests that the model accurately captures that particular feature of the model. Otherwise, it suggests a lack of fit for that statistic. In this section consider a single data set with two replicates per individual and fit the model SMEMN. For brevity, we only present the results we obtained when we generated errors from the skewed distribution.

Running 3 chains of 12,000 iterations with 2000 burn in iterations, results in Gelman-Rubin diagnostics of less than 1.03 for all parameters. Trace plots for parameters show no indications of non-convergence. Due to adaptive tuning during burn in, the acceptance rates for the latent variables hovers between 17% and 40% for the 300 individuals.

Table 2.6 give posterior quantiles for this model as well as for the truth. These regression coefficients associated with the linear component of the mean function have straightforward interpretation. For example, $\gamma_{1,ee}$ represents the change in reported EE, all else held equal, for males. In our example, the true value is 300 kcal, which means that for two persons who are identical except for their sex, the inexpensive measurement tool is biased such that it will record EE 300 kcal higher for the male than for the female. The parameter $\gamma_{2,ee}$ represents the effect of BMI on the device, and $\gamma_{3,ee}$ represents the effect of age.

Figure 2.11 shows plots of posterior predictive assessment for the latent variables when we generated skewed errors. Model SMEMN is less flexible than the model with the Dirichlet Process prior. The diagnostics look good given that the true variables are non-normal. This result is encouraging because we were concerned about the performance of a single bivariate normal distribution versus the Dirichlet Process as a prior for the bivariate latent vector.

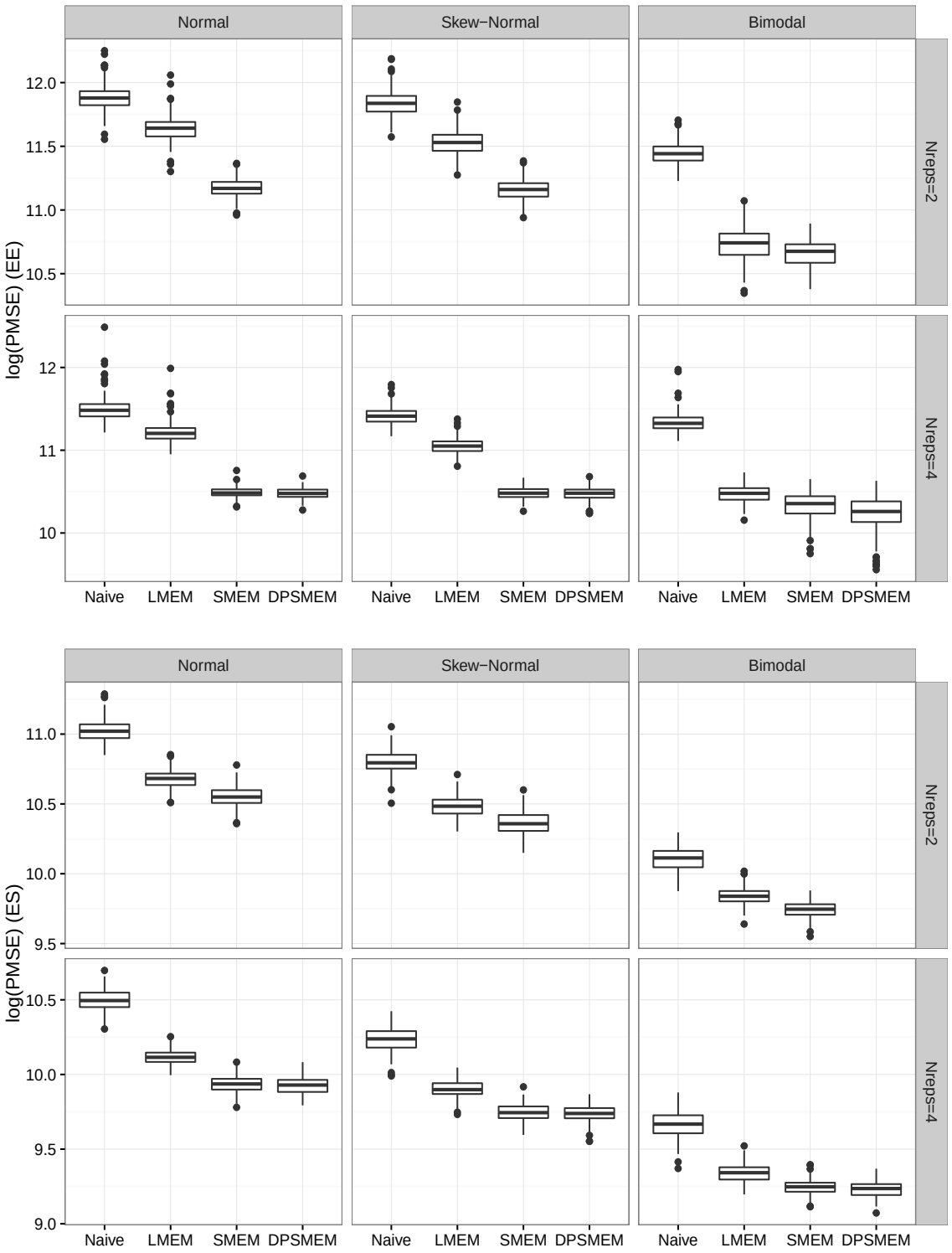

Figure 2.10: Log PMSE for EE Regression (top) and ΔES Regression (bottom) faceted by measurement error distribution and number of replicates

Figure 2.12 shows the same plots but for gold standard measurements. Even when the errors are skewed, the model seems to fit the data well because the true statistics are captured by the distribution of the statistic generated from the replicated datasets. Figure 2.13 shows the same plots, but now for the inexpensive noisy measurements. The model does not appear to be capturing the minimum value of EE and ΔES . Otherwise, the remaining plots look good, again even though the errors were generated from a skewed distribution. This suggests robustness of the approach to misspecification of the measurement error distribution.

Figure 2.14 shows the fitted spline between the values of EE and ΔES and the measurements obtained with the inexpensive measurement. The points correspond to the individual simulated data where the y value is the mean of the two replicates. The bold (red) line is the mean estimated spline function. We randomly selected 500 MCMC iteration draws for the spline, and plotted them behind the mean. Figure 2.15 gives the distribution of the number of knots for the spline for both the EE and ΔES splines. The splines are not overly complex typically use four or fewer knots.

Table 2.6: Parameter estimates for Model SMEMN for Skewed errors with correct specification of within person errors

	2.5%	25%	50%	75%	97.5%	True Values
σ_{yee}	368.55	384.27	392.91	402.25	422.86	422*
σ_{yes}	290.23	302.64	309.63	316.91	331.80	344*
σ_{wee}	237.48	249.19	255.82	262.64	276.41	250
σ_{wes}	64.86	67.77	69.41	71.11	74.52	72.862
$\gamma_{1,ee}$	268.58	294.32	308.09	321.41	347.99	300
$\gamma_{2,ee}$	6.00	8.38	9.64	10.89	13.28	14
$\gamma_{3,ee}$	-13.79	-11.69	-10.60	-9.54	-7.47	-7
$\gamma_{1,es}$	-256.60	-234.20	-222.84	-211.52	-188.25	-200
$\gamma_{2,es}$	4.32	6.70	7.92	9.19	11.60	8
$\gamma_{3,es}$	-7.18	-5.17	-4.15	-3.07	-1.00	-5

2.7 Calibration

The main goal of this work is to develop a calibration approach to "correct" the measurements of EE ΔES obtained with inexpensive, noisy measurements. That is, given a measurement of EE

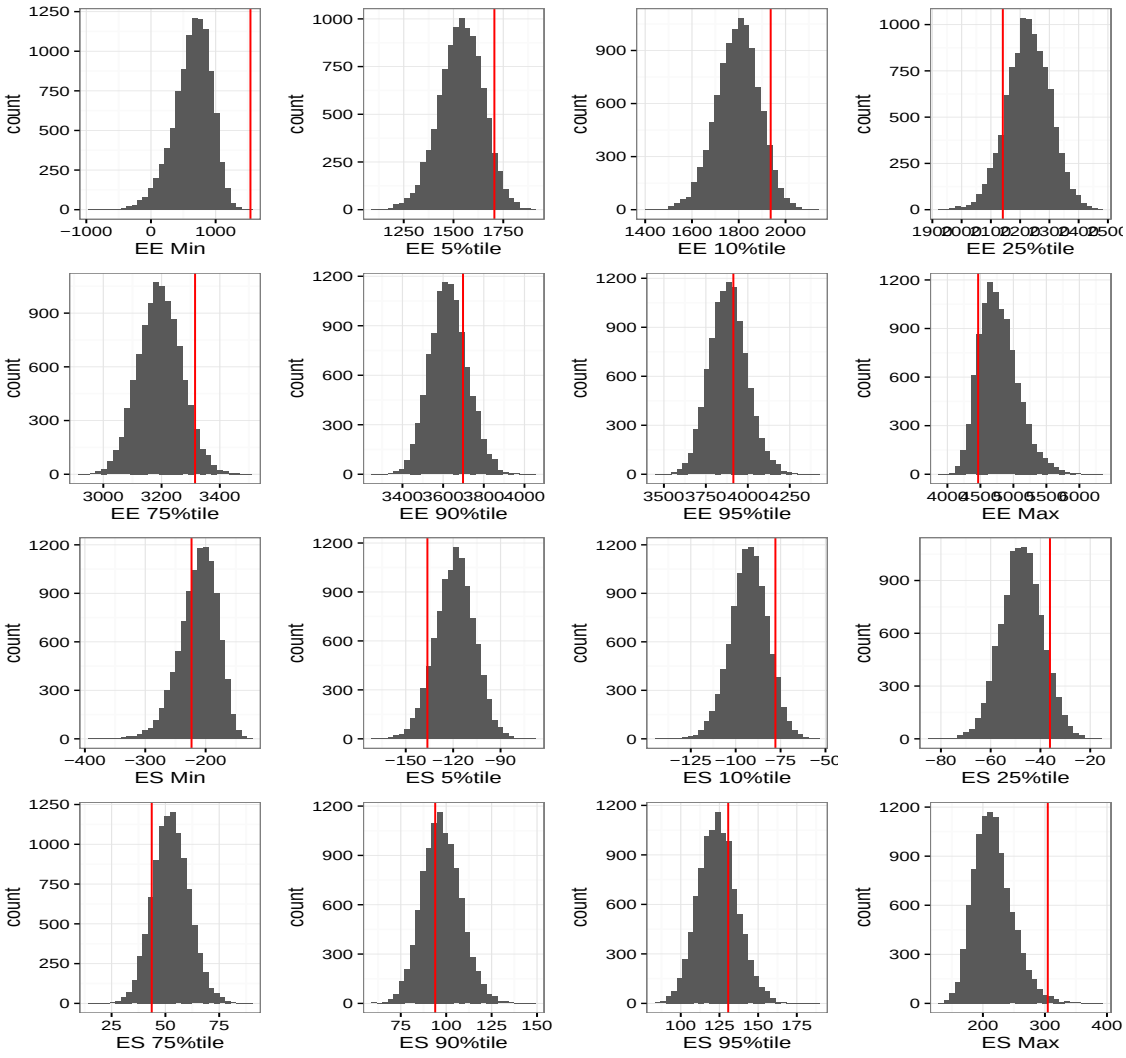

Figure 2.11: Posterior Predictive Discrepancy Measures For X^{EE} and $X^{\Delta ES}$ for Model SMEMN with Skewed Normal Errors

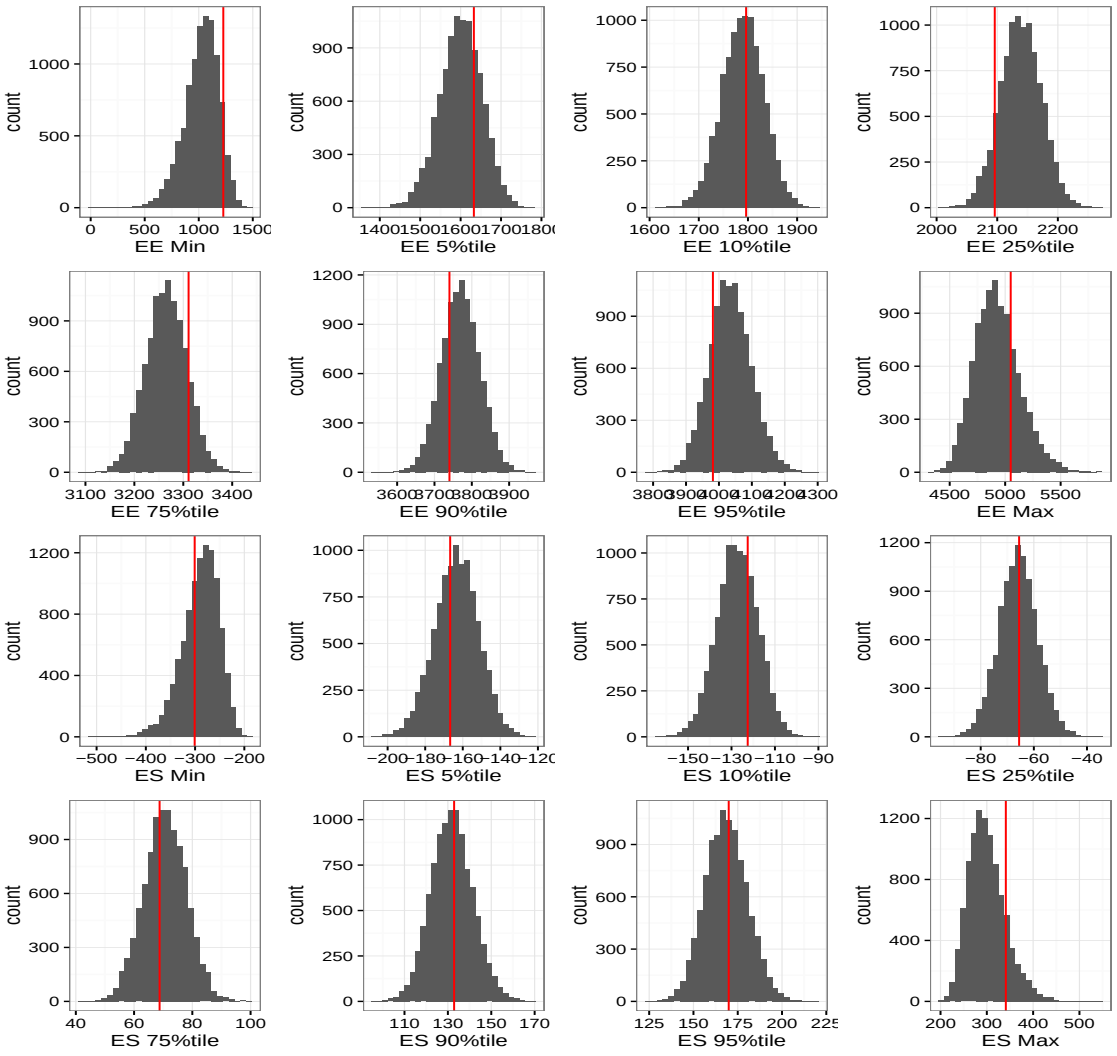

Figure 2.12: Posterior Predictive Discrepancy Measures For W^{EE} and $W^{\Delta ES}$ for Model SMEMN with Skewed Normal Errors

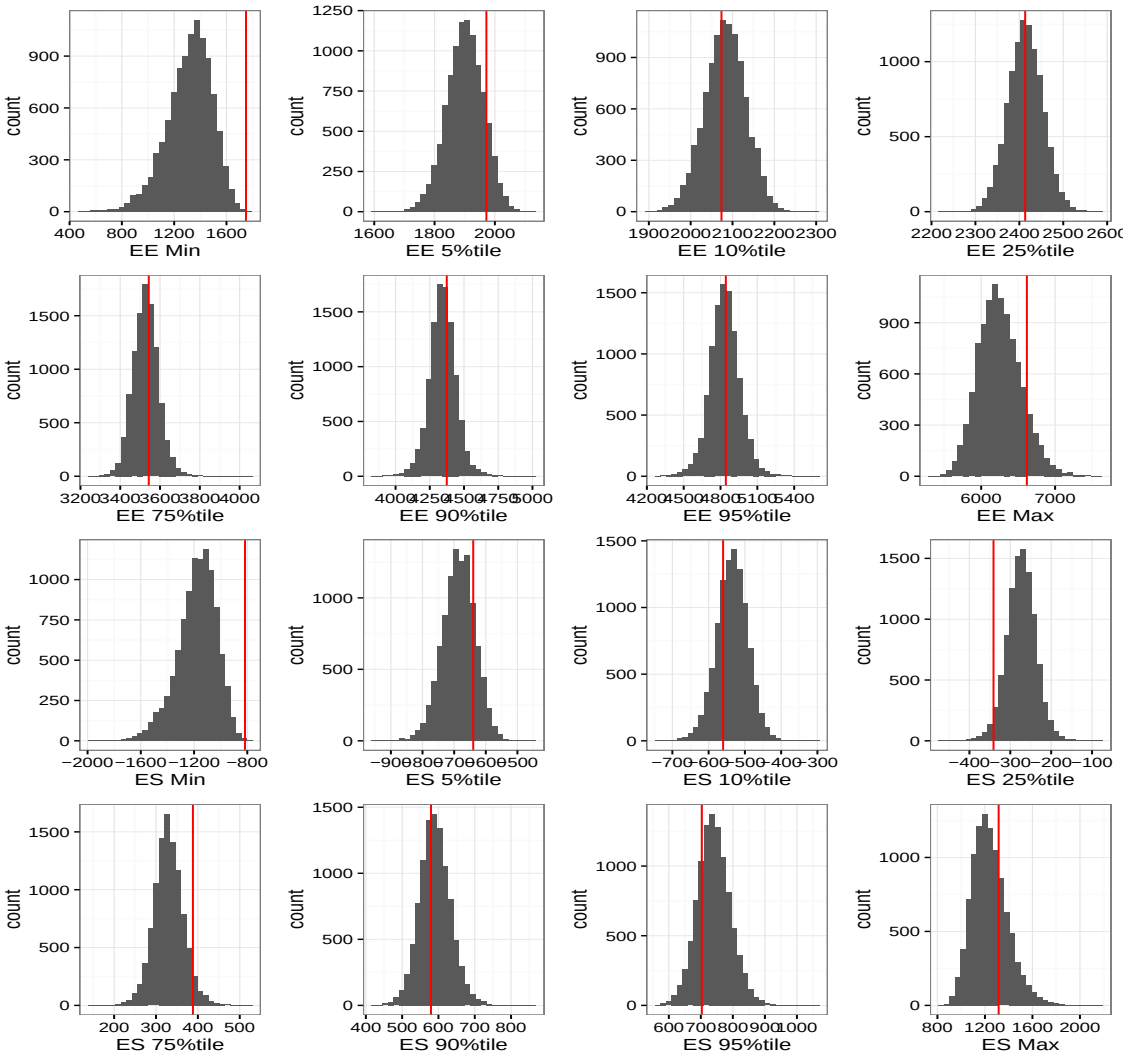

Figure 2.13: Posterior Predictive Discrepancy Measures For Y^{EE} and $Y^{\Delta ES}$ for Model SMEMN with Skewed Normal Errors

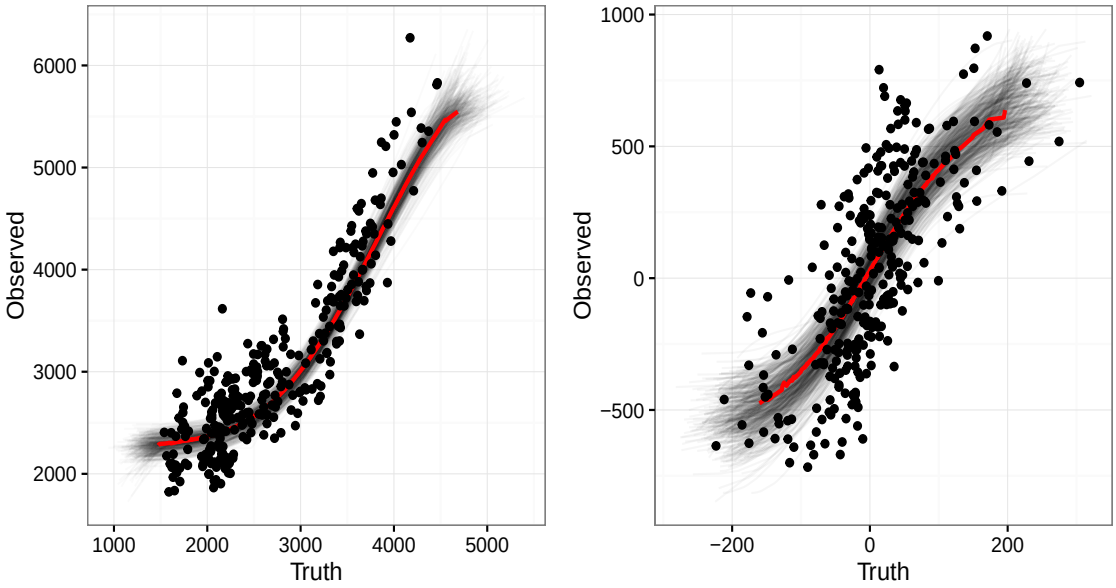

Figure 2.14: Spline function for Model SMEMN with Skewed Errors

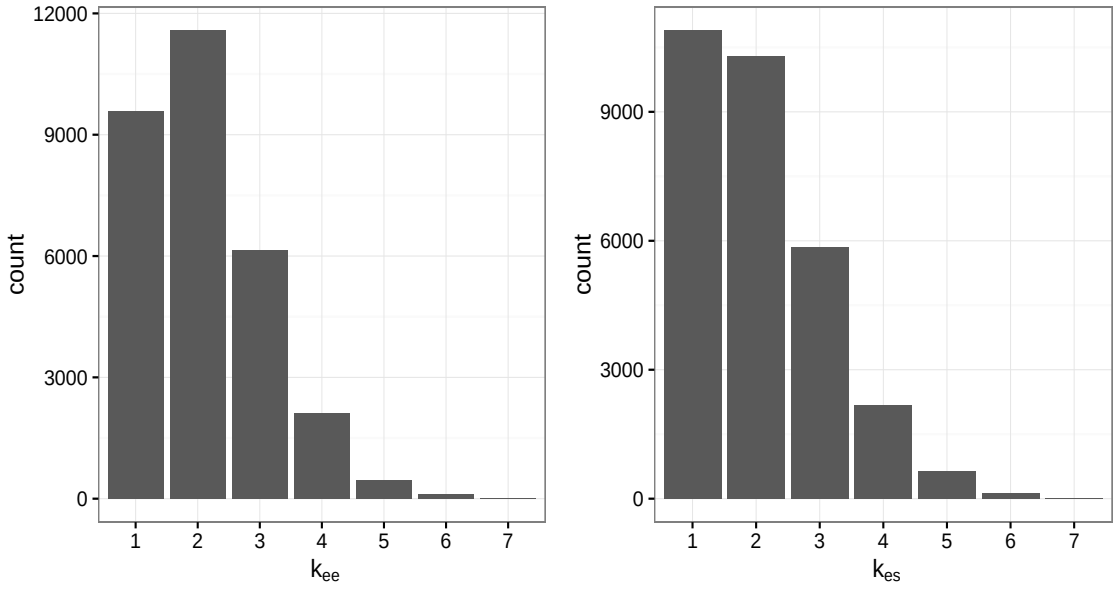

Figure 2.15: Number of knots for Model SMEMN with Skewed Errors

or ΔES from an inexpensive instrument and some demographic information, we can return a better estimate of the true value as well as a credible interval that shows the uncertainty in the estimate. Calibration for our models simply amounts to finding the inverse of the fitted models in (2.4), (2.5) as a function of Y instead of X, and Z are still measurement error free covariates. For a given observed value of Y and Z, and an estimate of $\boldsymbol{\gamma}$, the calibration for X is:

$$X_{calibrated} = s^{-1}(y - \boldsymbol{\gamma}' Z). \tag{2.80}$$

We cannot find the inverse in 2.80 in closed form so we find it numerically instead. To do so, we use `optimize` in `R` for the function $|s(x) - y^*|$ where $s()$ represents the regression function and y^* is the observed inexpensive measurement minus the vector of coefficients $\boldsymbol{\gamma}$ multiplied by the individuals' covariate values Z. The algorithm for our calibration for individual i is as follows:

For r = 1,...R

1. Calculate $y_i^* = y_i - \boldsymbol{\gamma}^{(r)'} Z_i$, where Z_i are the covariate values for individual i.

2. Use `optimize` for the function $|s_i(x) - y_i^*|$ to choose the value of x that will minimize the criterion, call this $x_{i,calibrated}$. Here, $s_i(x)$ is the predicted value of y_i for the given value x using the MCMC draw for the spline coefficients $\boldsymbol{\beta}_{\cdot}^{(r)}$, latent variables $(X^{EE(r)}, X^{\Delta ES(r)})$, and knot locations $(r_{ee}^{(r)}, r_{es}^{(r)})$ from the r^{th} draw of the chain.

As an example, suppose that we wish to calibrate three noisy measurements each from a different individual using Model SMEMN. We randomly select 3 individuals from the same data set used earlier to give results for model SMEMN. Individual 1 is male, BMI of 28.6, age 20.5; individual 2 is female, BMI of 21.5, age 30.1 and individual 3 is male, BMI 38.6 and age 22.8. Observed inexpensive measurements for these individuals, their true values, as well as 95% credible intervals for their mean calibrated truth under skewed normal errors are given in Table 2.7. Figures 2.16 shows histograms of 1000 calibrated draws for each individual for EE and ΔES measurements under skewed errors. Looking at the table and figure, one can see that the calibration helps pull the inexpensive measurement closer to the truth. In all cases, the calibration helped to improve

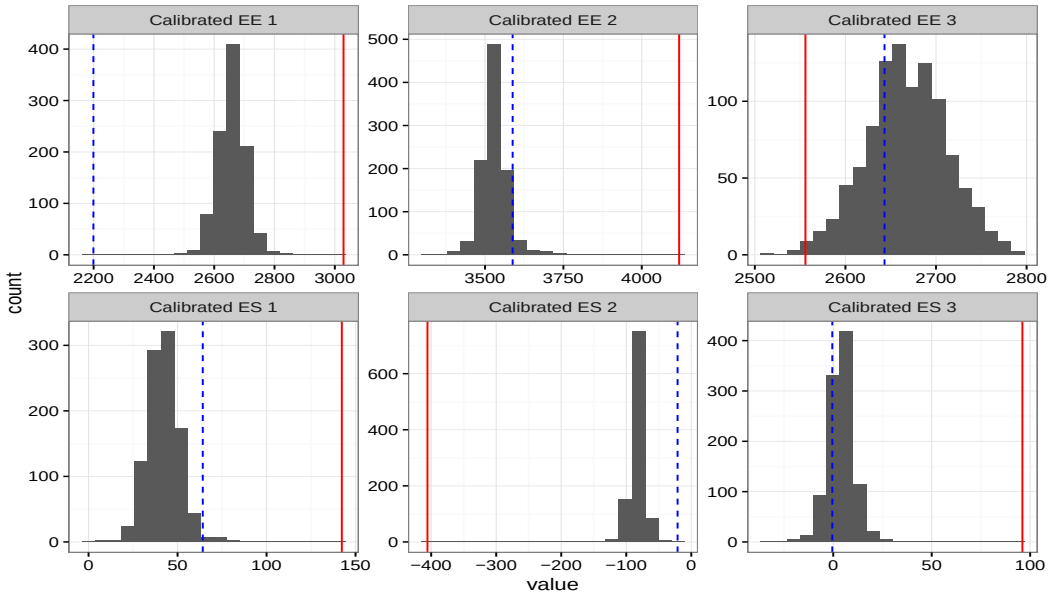

Figure 2.16: Posteriors of calibrated observations. Solid vertical line shows observed value from inexpensive measurement and dashed vertical line shows truth.

the estimate obtained from the inexpensive measurement. Running this on many of the simulated individuals had similar results.

Person		Lower	Median	Upper	Observed	True Value
1	EE	2574.18	2666.00	2736.39	3028.89	2199.25
2		3452.51	3525.18	3619.08	4119.26	3588.12
3		2571.99	2665.46	2744.65	2555.86	2643.14
1	ΔES	25.15	42.35	60.57	142.30	64.17
2		-104.21	-82.93	-63.90	-405.74	-21.08
3		-8.41	3.91	17.83	96.06	-0.48

Table 2.7: 95% credible interval for calibration estimate for inexpensive measurements for Skewed Errors

2.8 Discussion

In this chapter we presented a semi-parametric approach to model energy balance via its components EE and ΔES. We assume that we have gold standards for both quantities that are unbiased,

as well as inexpensive instruments that result in biased measurements of the truth. We propose a model where the form of the association between the unbiased and the biased measurements of EE (or of ΔES) is left unspecified and uses splines to estimate that function. This allows a flexible relationship between an inexpensive measurement and its unobserved truth. We assumed normality for the measurement error distributions which was justified through the use of diagnostics recommended by Caroll et al. (2006). We assumed that the gold standard measurements and inexpensive measurements are conditionally independent given the latent vector $(X^{EE}, X^{\Delta ES})$. We modeled the latent vector $(X^{EE}, X^{\Delta ES})$ using a bivariate normal distribution and a Dirichlet process. Although the Dirichlet process is more flexible and based on a weaker assumption, it required more replicate observations (mainly on gold standard measurements) than is feasible in practice in order to give stable results. The normality assumption was robust and resulted in stable and surprisingly reasonable results given the true structure of the latent variables. Because this model produced accurate estimates even with only two replicates of gold standard measurements per person, we believe that it is a plausibly useful model for this specific application unless more than two replicates per person are available. The resulting estimates and PMSE show the approach what we propose outperforms a simpler linear measurement error model and a naïve model that does not take measurement error into consideration.

The main motivation for constructing this model was to account for the error and bias in easy to administer measurements in order to calibrate inexpensive observations. We presented a simple way to do this calibration given an inexpensive measurement for EE and ΔES and values of gender, BMI, and age. Using a Bayesian approach we are easily able to get a posterior distribution for the mean calibrated estimate which also provides a measure of uncertainty. Our example shows that in some cases this calibrated estimate can be an improvement compared to the observed inexpensive measurement.

An extension of the model might allow for non-constant variance of the measurement error. At the time, we have limited data from a limited number of EE and ΔES measurement tools, but it is possible that non-constant variance will be an issue for some variables. Revisiting the conditional

independence assumptions might lead to improved predictions. We performed the same analysis assuming complete independence of EE and ΔES but found increased uncertainty in estimates, so we anticipate that modeling more of the dependence would be helpful. We assumed normality for the measurement errors which based on diagnostics from the EBS, assumed like a reasonable choice, but this is another model attribute that could be further explored.

CHAPTER 3. A BAYESIAN TWO-PART MODEL WITH MEASUREMENT ERROR: ASSESSING ADULT PHYSICAL ACTIVITY AND COMPLIANCE WITH 2008 CDC GUIDELINES

3.1 Introduction

The fact that physical activity plays a critical role in a healthy lifestyle is undisputed. Regular physical activity has been linked to prevention and treatment of cardiovascular disease (ODPHP (2017), Warburton et al. (2006),Reiner et al. (2013)), diabetes (ODPHP (2017), Warburton et al. (2006), Reiner et al. (2013)), cancer (ODPHP (2017), Warburton et al. (2006)), hypertension (Warburton et al. (2006)), osteoporosis (ODPHP (2017), Warburton et al. (2006)), depression (ODPHP (2017), Warburton et al. (2006)), obesity (Warburton et al. (2006), Reiner et al. (2013)), and Alzheimer's (Reiner et al. (2013)).

In recent years, and partially due to the obesity epidemic in the United States and elsewhere, the pace of research in physical activity and its effect on health has accelerated. According to the Centers for Disease Control and Prevention (CDC) over 70% of Americans age 20 and over are overweight or obese, and almost 40% are obese CDC (2017). These rates are not likely to decrease in the near future, since over 20% of teenagers are obese and 9% of children between 2 and 5 years old are already obese CDC (2017). In 2008, the Department of Health and Human Services issued the *2008 Physical Activity Guidelines for Americans* (PAG) which can be found at https://health.gov/paguidelines/guidelines/. This was the first time that physical activity guidelines were published by the federal government.

The PAG recommends that adults spend at least 150 minutes each week carrying out moderate-intensity activity, or at least 75 minutes in vigorous-intensity activity, or in some equivalent combination of moderate to vigorous physical activity (MVPA). Furthermore, they recommend that this activity be in intervals or *bouts* of at least 10 minutes. They define moderate-intensity to be

a 5 or 6 on a scale of 0 to 10, which as an example is the amount of energy that would be used during brisk walking. Vigorous-intensity is a 7 or 8 on the same scale, and jogging or lap swimming are examples of this level of activity. We provide more precise definitions later in this paper. In addition, the PAG recommends doing muscle-strengthening activities that involve all major muscle groups twice or more per week. These are the minimum levels of activity that are expected to have an effect on health. The report goes further and advises that any physical activity above the minimum will result in additional health benefits.

A public health question of interest is whether Americans of different ages adhere to these guidelines and if so, what proportion do so. This type of information is important for policy makers not only to assess compliance but also to design interventions that target certain subpopulations. Yet, there is no agreement about how to measure physical activity. Furthermore, the only nationally representative source of physical activity measurements is NHANES (CDC ()), which, as we argue later, used an instrument that tends to *underestimate* physical activity. To better understand the measurement error associated with different instruments, Iowa State University conducted an NIH-funded study to collect physical activity information (Beyler et al. (2015)). The objectives of the study were to understand the measurement error of three different instruments to measure physical activities in adults. We provide details about the study in the next section. Here, we use the measurements obtained using the Sensewear armband to develop a Bayesian two-part modeling approach that accounts for measurement error in physical activity observations and that can be used to estimate the time adults spend in MVPA in bouts of at least 10-minute duration. We compare our results with those that can be obtained by analyzing the NHANES data and discuss potential reasons for the differences we observed.

This paper is organized as follows. In Section 3.2 we describe the Physical Activity Measurement Survey (PAMS), explore the PAMS data, and review the literature on approaches to measure physical activity. In the next section, we develop a two-part model to jointly describe the distribution of daily number of 10-minute bouts and number of what we denote "excess minutes" in MVPA. In Section 3.4 we fit the model to the PAMS data, and in Section 3.5 we explain results in

the context of the application. In this section, we also compare results with those obtained using NHANES data and discuss differences. Model diagnostics are described in Section 3.6. Finally, we offer additional discussion and suggest future work in the Discussion section.

3.2 The Physical Activity Measurement Survey (PAMS)

3.2.1 Measuring physical activity – brief review

The term "physical activity" is not well defined. It is hardly surprising that many different approaches to quantify physical activity have been proposed in recent years. If we think of physical activity as the amount of energy expended per day during a short period (e.g., two weeks), then doubly-labeled water is considered to be the gold standard among measurement instruments (Bouten et al. (1996), Thomas et al. (2011b), Hall and Chow (2011)). However, it is impractical to use doubly-labeled water in large studies, not only because of cost but also because of respondent burden.

In practice, instruments such as accelerometers that measure movement have become commonplace. Accelerometers provide estimates of movement through uni-,bi-, or triaxial measurements. Measurements of activity are then often reported as "counts" (for Actigraph) or METs for SenseWear. Typically, raw accelerometer data are converted to counts or METs using proprietary algorithms, but there is growing interest in using the raw accelerometry data to create new, non-proprietary measures that attempt to capture physical activity more fully (Bai et al. (2014)). Urbanek et al. (2017) uses the full, raw accelerometry data to create new measures of stride-to-stride gait variation. He et al. (2014) uses movelets (Bai et al. (2012)) to classify activity types based on accelerometry data. This may help compensate for the fact that accelerometry data provide no information about *the context* in which physical activity takes place. There is a rich literature that focuses on the relationship between total counts per day and some outcome variable (Schrack et al. (2014), Steeves et al. (2015a)). Other authors use count data at the hour level to further understand how physical activity levels vary by demographic groups (Steeves et al. (2015a), Steeves et al. (2015b)). Functional data analysis (FDA) is also popular to model and analyze high-frequency, accelerometer

data. Xiao et al. (2015) provides methods to model the systematic and random patterns of physical activity while accounting for dependence on covariates such as age and gender. Fan et al. (2015) using functional ANOVA to assess the circadian activity profiles of teenage girls. Goldsmith et al. (2016) uses functional scalar regression to understand the association between physical activity and a variety of covariates. FDA permits understanding how relationships between physical activity and covariates evolve during a day. These relationships can then be used to tailor interventions such as after-school programs for kids whose parents work full time.

3.2.2 Description of the data

PAMS was a survey conducted over two years starting in 2009 in Iowa. The goal was to obtain information on physical activity of adult men and women. The survey was a two-stage survey fielded in four counties and included two strata per county. In each county there was a "high minority population" and "low minority population" stratum to improve chances of recruiting African American and Hispanic individuals. Eligible participants included adults between 21 and 71, with the ability to engage in physical activity, who were not pregnant or lactating, were able to speak English or Spanish, and had a landline in their place of residence. A summary of the demographic characteristics of PAMS participants is given in Table 3.1.

Individuals provided energy expenditure (EE) information on two separate occasions using two different measurement tools. Here we focus on EE data collected using a SenseWear armband (SWA). In order to mitigate dependence in activity across days for an individual, the two measurements were taken 2-3 weeks apart. The SWA provides minute by minute information on MET-minutes. The method with which the SWA calculates MET-minutes is proprietary, but SWA's measurement properties and validity have been studied. Hills et al. (2014) and Hill et al. (2010) found the SWA to be a reasonably accurate measure of physical activity. Santos-Lozano et al. (2017) and Scheers et al. (2013) found that the SWA tends to overestimate MVPA. Calabro et al. (2014) also found the SWA to slightly overestimate physical activity, but it was much closer to truth than the other accelerometers used which underestimated physical activity by a greater magnitude.

Casiraghi et al. (2013) notes that the SWA is a good measurement for certain activities like running and walking, but its use is limited in activites like cycling and swimming. A Metabolic Equivalent of Task (MET) is a measure of energy cost for a particular physical activity. Formally, 1 MET is defined as 0.0175 kcal/kg/min expended. METs can be thought of as a multiplicative effort to carry out the activity relative to resting state. An activity that is classified as 5 METs then requires about 5 times the energy that is required to be at rest. The *Compendium of Physical Activities* gives MET values for many common daily activities (Ainsworth et al. (2011)). MET-minutes are the number of minutes in an activity multiplied by the MET value of that activity.

Table 3.1: Characteristics summaries of PAMS.

	Male	Female
Count	610	848
Age Range	21-70	21-71
Mean Age	48.21	51.32
SD Age	13.07	11.88
BMI Range	14.41-62.95	15.41-72.88
Mean BMI	30.31	30.87
SD BMI	6.53	8.24
No. of African Americans	46	78
No. of Hispanics	26	34
Smoker	128	150
College Graduate	248	305
Physical Job	182	62

The *Physical Activity Guidelines* (PAG) defines activities with METs ranging from 3.0 to 6.0 as moderate intensity and activity with METs greater than 6.0 as vigorous. This means that the recommendation of 150 minutes of moderate physical activity is equivalent to $150 \times 3.0 = 450$ MET-minutes per week. This MET-min framework is convenient because the guidelines allow for combinations of moderate and vigorous activity. If we know the MET values of the person's activities it is possible to calculate the total MET-minutes for the week and compare the result to the recommendation of 450 MET-minutes in the PAG. Another stipulation in the *PAG* is that activity must occur in bouts of at least 10 minutes to count toward this total. In practice, what

constitutes a *bout* is less clear. If an individual is jogging for eight minutes, then has to wait at a stop light for a minute, then continues to jog for 3 minutes, does this count as at least 10 minutes of at least moderate activity? To address the research questions we need an operational definition of what constitutes a *bout*.

3.2.3 Definition of *bouts*

To determine what constitutes a *bout*, we follow the commonly used definition that establishes that 8 out of 10 minutes in at least 3 METs (Tucker et al. (2011), Kim et al. (2015)) is a bout. This means that at least 8 minutes out of 10 minutes must be in at least moderate physical activity to count toward the recommended guidelines. We allow the 8 out of 10 to move along a rolling window, by shifting a 10 minute window, minute by minute, to determine if the time counts. As long as we observe ≤ 2 minutes in less than moderate activity (<3 METs), the "clock continues to count" minutes for that bout. Once we observe ≥ 3 minutes in less than moderate activity, the "clock stops" at the minute before the 3rd minutes is reached. Further, we do not allow the final 2 minutes of activity to be below moderate level. For example, someone who is in moderate intensity for 8 minutes, then is sedentary for the next 7 minutes, achieves 0 minutes toward the weekly total. Then for each trial for each individual, we calculate the number of bouts during the 24 hour period as well as the total number of MET-mins during the respective bouts. Figure 3.1 shows the distribution of the number of bouts and of the total MET-mins in a 24 hour period for all individuals and trials. In about 24% of 24 hour periods, individuals participated in zero bouts of MVPA, meaning that 0 MET-mins are counted toward the recommended activity times. About 11% of individuals in PAMS had zero bouts on both study days.

Total MET-mins in MVPA in at least 10 minute bouts is zero for individuals with zero bouts, and is $\geq$30 for individuals with a minimum of one bout at 3 METs for 10 minutes (10 minutes $\times$ 3 METs = 30 MET-mins in MVPA). To account for these constraints, we define Y_{1ij} as the number of bouts for individual i during trial j and Y_{2ij} as the average MET-mins in MVPA per bout minus 30, set as zero if $Y_{1ij} = 0$. We refer to Y_{2ij} as the average excess MET-minutes. There

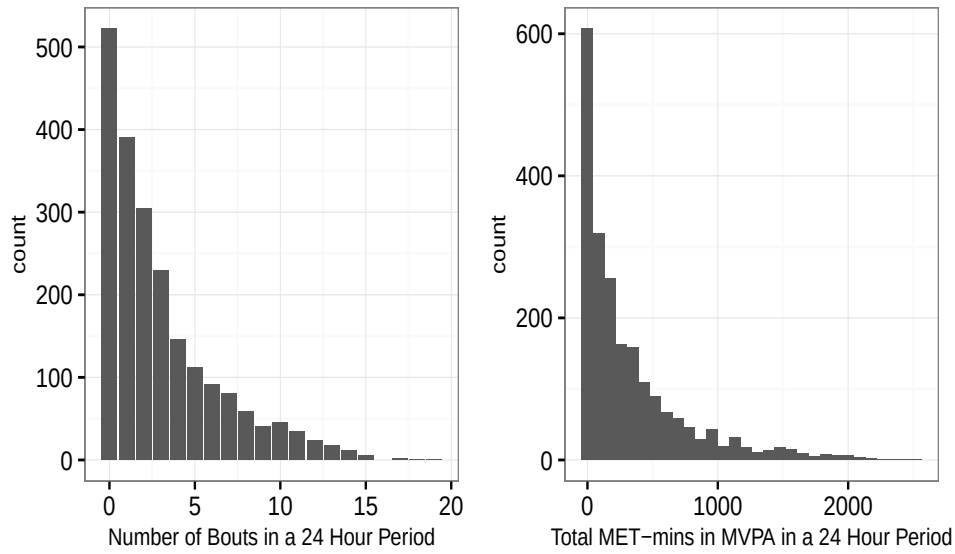

Figure 3.1: Left: Distribution of number of bouts in a 24 hour period for all individuals and trials. Right: Distribution of total MET-mins from bouts in 24 hour period for all individuals and days.

were several outliers in both number of bouts and total MET-minutes; we removed persons with more than 2500 total MET-min. Figure 3.2 plots Y_1 against Y_2. We restrict the range in the plot and limit the number of bouts to be between 1 and 13 per day to ensure that that there are at least 20 observations in each bout boxplot. Even after adjusting for the number of bouts, Y_{2ij} still seems to be positively related to number of bouts. We incorporate this dependence into the model that we discuss in Section 3.3.

3.2.4 Distribution of Y_2

The distribution of Y_2 is heavily right skewed, and a standard first step is to use transformations in an attempt to obtain a distribution that approximates a normal distribution. In physical activity and nutrition applications, the log transformation is often used. Using log-transformed the positive Y_2 values, we performed the linear regression:

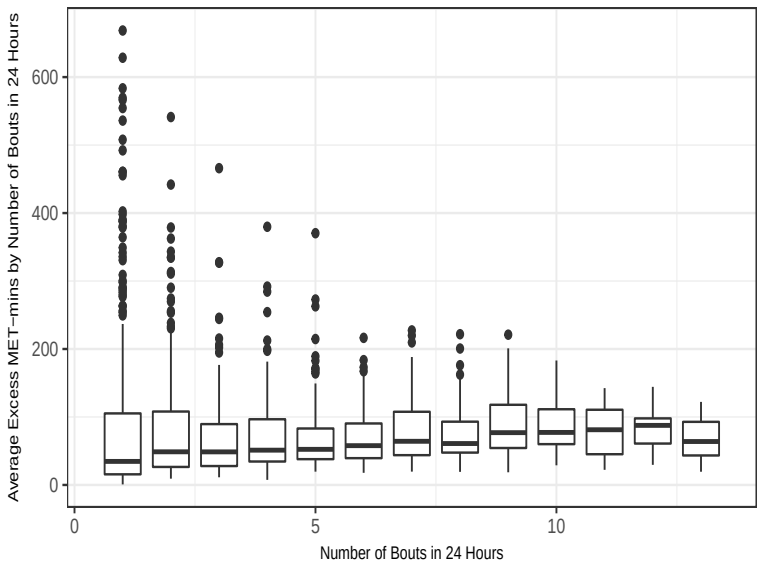

Figure 3.2: Average excess MET-mins per bout by number of bouts in 24 hour period for all individuals and days.

$$log(Y_{2ij}) = \boldsymbol{\theta}'\mathbf{Z_i} + e_{ij}, \tag{3.1}$$

$$e_{ij} \overset{iid}{\sim} N(0, \sigma^2). \tag{3.2}$$

Figure 3.3 shows a QQ plot for the residuals for the above model. A Shapiro-Wilk test for normality of the residuals results in a p-value of 0.15; this along with the QQ plot suggests that the empirical distribution of the log transformed data approximates a normal distribution, which allows us to use a lognormal distribution to model the Y_2 in the original scale.

3.2.5 Checking for day effect of observations

We check whether the two repeated measures on each subject can be assumed to be independent within an individual. If so, we can assume that conditional on individual, observations are exchangeable. Although we use the SWA to measure physical activity which is unlikely to exhibit response bias, and although the trials are separated by about two weeks, there is still reason to believe that individuals act differently on the first and second trials.

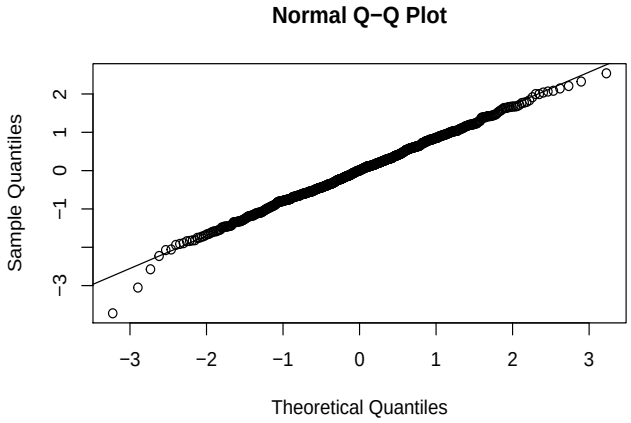

Figure 3.3: Normal quantile plot of residuals for log transformed Y_2 regression. This plot shows a $\log(Y_2)$ is approximately Normally distributed.

To check whether assuming exchangeability within individual in Y_1 is reasonable, we test whether the mean of paired differences $(Y_{1i1} - Y_{1i2})$ equals zero, and also explore the distribution of the differences. A t-test of the hypothesis of no difference in number of bouts within persons resulted in a statistic of 0.122 with a p-value of 0.24 indicating no evidence of a difference in the mean number of bouts on days one and two. To explore the joint distribution of Y_{1i1} and Y_{1i2}, we created a two way contingency table for number of bouts in day one versus number of bouts in day two including only individuals who had two observations. Figure 3.4 shows the frequency of individuals that had the particular combination of bouts on days one and two. This matrix is like a "transition matrix". Bowker (1948) proposed a test for symmetry in m by m contingency tables. The null hypothesis of Bowker's test is that $\pi_{lk} = \pi_{kl} \; \forall \; l \neq k$ where π_{ij} is the true frequency in the ijth cell. Using the function `mcnemar.test` in `R`, we tested the symmetry of the contingency table in Figure 3.4. The resulting chi-square statistic was 67.266 with 55 df and a p-value of 0.1241. This test is sensitive to the presence of zero or low counts, so we also implemented the same test on smaller subsets of the contingency table to ensure that the results were consistent. In all cases, we failed to reject the null hypothesis, which suggests that within individual measurements of number of bouts can be assumed to be exchangeable.

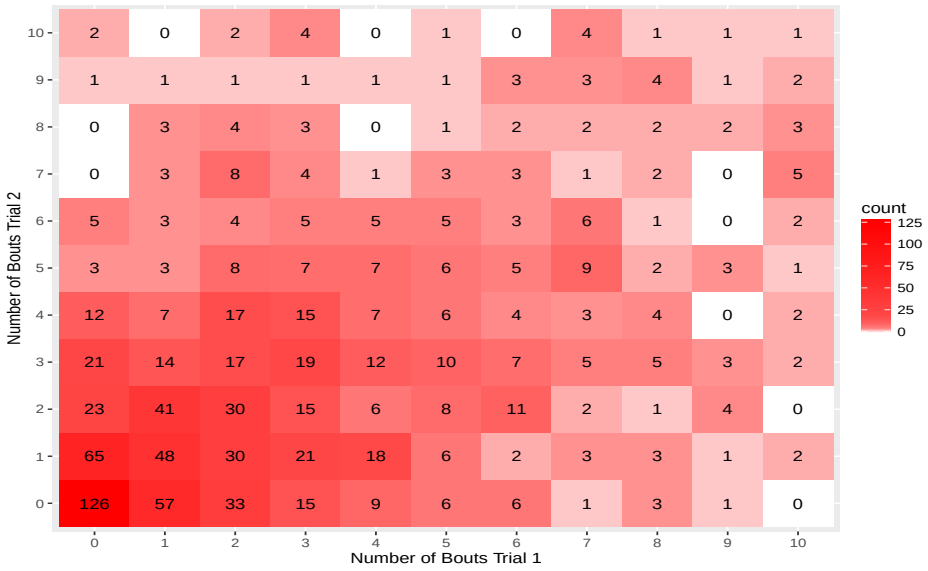

Figure 3.4: 2 Contingency way table for individuals with two observations based on number of bouts per trial. Truncated at 10 bouts due to sparsity beyond that.

We also need to check whether a day effect is present in Y_2, which depends on the number of bouts (Figure 3.2). Therefore we need to account for the effect of bouts when checking for day effect on Y_2. We are also interested in knowing whether there is an effect of weekend on Y_2. The relationship between Y_1 and Y_2 appears to be approximately linear; to explore the association between Y_2 across days fit we the following linear model:

$$\begin{aligned} Y_{2i1} - Y_{2i2} &= \beta_0 + \beta_1(Y_{1i1} - Y_{1i2}) + \beta_2(Weekend_{i1} - Weekend_{i2}) + \epsilon_i, i = 1, 2, ..., n \\ \epsilon &\overset{iid}{\sim} N(0, \sigma^2), \end{aligned} \tag{3.3}$$

where $Weekend_{ij}$ is an indicator for weekday (M-F) versus weekend (Sat or Sun), β_0 represents the day effect and β_2 represents weekend effect. Estimates of model parameters are given in Table 3.2. There does not appear to be either a day or a weekend effect on Y_2. We also checked for weekend effect on Y_1 using a paired t-test, which resulted in a p-value >0.5. Since there is no obvious indication of a day effect, we will assume that observations within individuals are exchangeable.

Table 3.2: Results from performing OLS on the model in 3.3. Since the coefficients for β_0 and β_2 have large standard errors relative to their estimates, there is no evidence that observations are not exchangeable within individual or of a weekend effect.

Coefficient	Estimate	Std Error	P-value
β_0	1.245	2.621	0.635
β_1	3.748	0.781	<0.0001
β_2	-1.709	4.344	0.694

3.3 Model for MET-mins in MVPA During at least 10 Minute Bouts

To answer the questions presented in the introduction of this chapter, we need to model MET-mins in MVPA occurring in at least 10 minute bouts. The added challenge we face is that the observations are contaminated by measurement error. We introduced the correlated random variables Y_{1ij} and Y_{2ij} earlier as the response variables. In this section, we present a measurement error model for Y_1 and Y_2.

3.3.1 Notation and data

After removing outliers and individuals without a replicate observation, we have $N = 2114$ observations obtained on $n = 1057$ individuals. We let i represent individual, $i = 1, ..., 1057$ and j represent the measurement occasion, so $j = 1, 2$. We let $\mathbf{Y_{ij}} = (Y_{1ij}, Y_{2ij})$. Similarly, we let $\mathbf{Y_1} = \{Y_{1ij}\}_{\forall i,j}$ and $\mathbf{Y_2} = \{Y_{2ij}\}_{\forall i,j}$, and $\mathbf{Y} = (\mathbf{Y_1}, \mathbf{Y_2})$. We define a vector $\mathbf{Z_i}$ of dimension eight, that includes covariates for individual i: gender, age, indicators for Black, Hispanic, smoker, college degree, and physical job. The full model matrix is $\mathbf{Z} = (Z_1', Z_2', ..., Z_{1057}')$. There were 315 instances of item non-response for occupation in for our 1057 individuals, so we imputed the missing values using predictions from a logistic regression with physical job as the response and all remaining covariates in Z as covariates. We use T_{1ij} to denote individual i's unobservable true number of bouts on day j and T_{2ij} is individual i's unobservable true average excess MET-minutes per bout on day j. We let t_{1i} and t_{2i} be the expected values of T_{1ij} and T_{2ij} conditional on individual i, respectively. We refer to these quantities as individual i's *usual* number of bouts in a day and *usual*

average excess MET-mins per bout, respectively. More formally:

$$\begin{aligned} t_{1i} &\equiv E(T_{1ij}|i), \\ t_{2i} &\equiv E(T_{2ij}|i). \end{aligned} \tag{3.4}$$

Following Kipnis et al. (2009) we adopt a classical measurement error formulation and assume that the measurements of physical activity are unbiased for the usual activity levels. This is likely to be a plausible assumption because the measurements are obtained using an objective instrument, but there could still be systematic bias in the armbands. We also assume that the armband records zero bouts if and only if individual i participated in zero bouts of activity on day j. Formally, these assumptions can be expressed as:

$$\begin{aligned} t_{1i} &= E(Y_{1ij}|i), \\ t_{2i} &= E(Y_{2ij}|Y_{2ij} > 0, i) \times P(Y_{2ij} > 0|i), \\ P(T_{1ij} > 0|i) &= P(Y_{1ij} > 0|i). \end{aligned} \tag{3.5}$$

By construction, $P(Y_{2ij} > 0|i) = P(Y_{1ij} > 0|i)$. In the end, we are interested in an individuals' *usual* total MET-minutes in MVPA for a day, which is defined as:

$$t_{3i} \equiv 30t_{1i} + t_{2i} \times t_{1i}. \tag{3.6}$$

3.3.2 Modeling number of bouts

We first consider the observed number of bouts Y_{1ij}. Figure 3.1 shows the distribution of all Y_{1ij}. The number of bouts Y_1 is a count, so a natural place to start is the Poisson distribution. We add a random intercept in the mean for individual effects, which accommodates potential overdispersion between individuals. However, we also need to account for overdispersion within individuals. Under a Poisson model, conditional on an individual, the mean and the variance are equal for that individual. We fit a Poisson model to the Y_1 data, similar to that in 3.7, and simulated data from the fitted model. Figure 3.5 shows the distribution of mean within person

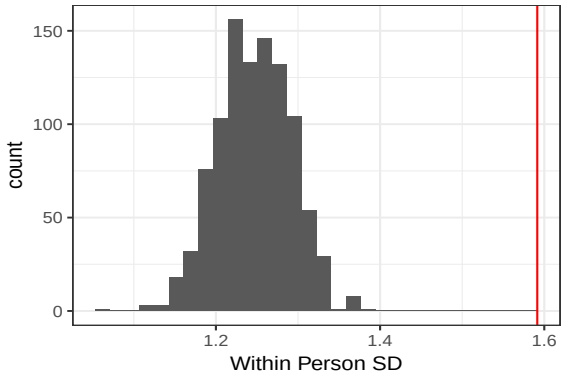

Figure 3.5: Posterior predictive model assessment for Poisson model. Statistic is mean within person standard deviation. Vertical line indicates observed value.

standard deviations for each simulated dataset as well as the truth as a vertical line. This shows the standard Poisson distribution is not sufficient for these data.

An alternative to the Poisson distribution that allows for a more flexible mean-variance relationship is the Generalized Poisson distribution (Consul (1988)). The Generalized Poisson distribution is indexed by two parameters, θ and λ. The Generalized Poisson is overdispersed relative to a Poisson distribution if $\lambda > 0$, underdispersed if $\lambda < 0$ and a regular Poisson if $\lambda = 0$. Because it can allow for underdispersion, the Generalized Poisson distribution is more flexible than the negative Binomial distribution. When only interested in the overdispersed case, it is common to consider the case where $0 < \lambda < 1$ (Czado et al. (2007), Joe and Zhu (2005), Scollnik (1998)). When $0 < \lambda < 1$, the probability mass function and first two moments of the distribution can be written directly, ie. without truncation and normalization (Consul (1988), Scollnik (1998)). In this case, the expected value of Y_1 is $\frac{\theta}{1-\lambda}$ and the variance is $\frac{\theta}{(1-\lambda)^3}$. Reparamaterizing the distribution in terms of the mean, μ, the variance is $\frac{\mu}{(1-\lambda)^2}$. At this point we only concern ourselves with overdispersion, thus the restriction that $0 < \lambda < 1$ is appropriate.

We model the mean of the Generalized Poisson distribution as a function of covariates: age, gender, smoking indicator, education level, race, and occupation as well as the individual random effect. We assume joint normality for these random effects, and their complete specification is given

in the next section. We chose the priors for λ and $\boldsymbol{\gamma}$ to be proper and independent, but relatively non-informative. Our model for Y_{1ij} is written as:

$$\begin{aligned} Y_{1ij}|b_{1i}, Z_i &\overset{ind}{\sim} \text{GenPoisson}(\mu_{1i}, \lambda) \\ \mu_{1i} &= E(Y_{1ij}|i) = e^{Z_i'\boldsymbol{\gamma}+b_{1i}} \\ \lambda &\sim \text{Uniform}(0,1) \\ \boldsymbol{\gamma} &\sim N\left(\mathbf{0_8}, 100 I_{8\times 8}\right). \end{aligned} \tag{3.7}$$

3.3.3 Modeling total excess MET-minutes

Figure 3.3 showed that a log-Normal distribution appears to fit the positive values of Y_2 nicely. However, recall that there are a large number of Y_2 observations that equal zero, and the rest are positive and continuous. This type of data is commonly referred to as "semicontinuous data" and occurs often in the fields of epidemiology and nutrition. Many models for semicontinous data have built upon the work of Olsen and Schafer (2001) and soon thereafter Tooze et al. (2002). Neelon et al. (2011) and Neelon et al. (2015) propose Bayesian approaches for estimation in these models. Good reviews on inference in semicontinous data are Neelon et al. (2016a) and Neelon et al. (2016b). Tooze et al. (2006) propose a model to estimate distributions of episodically consumed foods where the observed distribution also has a peak at zero. Kipnis et al. (2009) and Kipnis et al. (2016) propose a measurement error approach for semicontinous data via regression calibration in the context of a nutrition application. Zhang et al. (2011b) and Zhang et al. (2011a) discuss applications and extensions to the multivariate version of this measurement error model .

We have already assumed that individuals are independent, and we also argued in favor of exchangeability of observations within individuals. To account for measurement error and the large number of zeros in the sample we propose the following model:

$$
\begin{aligned}
Y_{2ij}|b_{2i}, Z_{1i} &\overset{ind}{\sim} (1-\pi_i)I(Y_{2ij}=0) + \pi_i \text{LogNormal}(\mu_{2i}, \sigma_y^2) I(Y_{2ij}>0), i=1,...,n, j=1,2, \\
\mu_{2i} &= E(logY_{2ij}|Y_{2ij}>0, i) = Z_i'\boldsymbol{\beta} + b_{2i}, \\
b_{1i}, b_{2i} &\overset{ind}{\sim} N\left(\begin{bmatrix}0\\0\end{bmatrix}, \Sigma_b\right), \\
\Sigma_b &\sim \text{Inverse-Wishart}(3, I_{2\times 2}) \\
\sigma_y^2 &\sim \text{Inverse-Gamma}(1,1), \\
\boldsymbol{\beta} &\sim N\left(\mathbf{0_8}, 100 I_{8\times 8}\right),
\end{aligned}
\tag{3.8}
$$

where $\pi_i = P(t_{2ij} > 0|i) = P(Y_{1ij} > 0|i)$ is individual i's probability of participating in at least one bout, which can be calculated using the Generalized Poisson probability mass function given parameters $\boldsymbol{\gamma}, \lambda, b_{1i}$ and design matrix Z. We chose the priors for $\boldsymbol{\beta}, \sigma_y^2, \Sigma_b$ to be conjugate and independent to ease the computational burden. Sensitivity analysis showed little effect of the priors on inference for the variance components.

3.3.4 Full model likelihood

Putting all these parts together and relying on the assumptions of exchangeability of observations within individuals, the full likelihood for an individual can be written as:

$$
\mathbf{L_i}(\boldsymbol{\theta}) = \int\int \prod_{j=1}^{2} f(Y_{2ij}|\boldsymbol{\theta}, b_{2i}, b_{1i}) f(Y_{1ij}|\boldsymbol{\theta}, b_{1i}) f(b_{1i}, b_{2i}|\boldsymbol{\theta}) db_{1i} db_{2i}, \tag{3.9}
$$

where $f(Y_{1ij}|\cdot)$, $f(Y_{2ij}|\cdot)$, and $f(b_{1i}, b_{2i}|\cdot)$ are as defined in (3.7) and (3.8). Along with the assumption of independence between individuals, the full likelihood is:

$$
\mathbf{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} \mathbf{L_i}(\boldsymbol{\theta}). \tag{3.10}
$$

3.3.5 Estimating distribution of usual daily MVPA

Our goal is to estimate the proportion of Americans who are in compliance with the *2008 Physical Activity Guidelines,* on average. To answer this question, we focus on the distribution of *usual* total MET-minutes in MVPA for individuals from a specified population in a day. We specify the population in which we are interested through the design matrix Z. To estimate this distribution, we take L draws from the posterior distribution of parameters $\boldsymbol{\theta}$ and simulate L sets of new person effects b_1, b_2, and calculate L corresponding values of t_3 using (3.5) and (3.6). The steps to estimate the distribution of usual daily MVPA for a specific population are the following:

For ℓ from $\ell = 1, 2, ..., L$ do:

1. Sample $\boldsymbol{\theta}^{(\ell)}$ from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{Y})$.
2. Simulate $b_{1i}^{(\ell)}$ and $b_{2i}^{(\ell)}$ from $p(b_{1i}, b_{2i}|\boldsymbol{\theta}^{(\ell)}, Z_i)$ for $i = 1, ..., n$.
3. Compute $E(Y_{1ij}|\boldsymbol{\theta}^{(\ell)}, Z_i, b_{1i}^{(\ell)}))$, $E(Y_{2ij}|Y_{2ij} > 0, \boldsymbol{\theta}^{(\ell)}, Z_i, b_{2i}^{(\ell)}, b_{1i}^{(\ell)})$, and $P(Y_{1ij} > 0|i)$ according to model specifications in 3.7, 3.8.
4. Compute $t_{3i}^{(\ell)} = 30E(Y_{1ij}|\boldsymbol{\theta}^{(\ell)}, Z_i, b_{1i}^{(\ell)}) + E(Y_{2ij}|Y_{2ij} > 0, \boldsymbol{\theta}^{(\ell)}, Z_i, b_{2i}^{(\ell)})$ $\times\ P(Y_{1ij} > 0|\boldsymbol{\theta}^{(\ell)}, Z_i, b_{1i}^{(\ell)}) \times E(Y_{1ij}|\boldsymbol{\theta}^{(\ell)}, Z_i, b_{1i}^{(\ell)}))$ for $i = 1, ..., n$.

The proportion of individuals from the population who meet the PAG in the ℓ^{th} draw is given by:

$$p^{(\ell)} = \frac{1}{n}\sum_{i=1}^{n} I\left(t_{3i}^{(\ell)} \geq \frac{450}{7}\right). \tag{3.11}$$

If there are weights w_i associated with the individuals of the design matrix Z, estimates of percentiles of the distribution of t_3 can be obtained unbiasedly. A weighted estimate of compliance for the population of interest can be calculated as:

$$p^{(\ell)}_{weighted} = \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} I\left(t^{(\ell)}_{3i} \geq \frac{450}{7}\right) w_i. \tag{3.12}$$

Recall that our model is estimating usual *daily* MET-minutes in bouts, and our model already considers how often individuals participate in at least a bout of MVPA. Because of this, we can consider weekly activity to be 7×usual daily MET-minutes in bouts.

3.4 Estimation

We proceed with estimation via MCMC as the joint posterior distribution is too complex to derive analytically. Neelon et al. (2011) and Neelon et al. (2015) propose Gibbs algorithms for two-part models for semicontinuous data that are nearly or completely conjugate. Measurement error adds a level of complexity to the sampler. We construct a Gibbs algorithm for drawing samples from the posterior distribution, and since many of the priors are not conjugate, we need to use a Metropolis-within-Gibbs sampler. The regression parameters $\boldsymbol{\beta}$ in the log-Normal part of the model, the random effects covariance matrix Σ_b, and the variance component of the log-Normal σ^2_y have conjugate priors and can be drawn directly from their full conditional distributions. For the regression parameters $\boldsymbol{\gamma}$ in the Generalized Poisson part and the random effects vectors for individuals (b_{1i}, b_{2i}), we use a random walk with a Normal proposal that is adaptively tuned during burnin. We sample λ using an independence Metropolis sampler, with a Beta proposal distribution. The Gibbs algorithm was written in C++ and R. Full conditional distributions can be found in the Appendix. To get starting values, we fit similar generalized linear models for Y_1 and Y_2 fit independently using `glm` in `R`. We used the resulting MLE's and lower and upper bound of 99.99% confidence intervals as starting values for the regression parameters. We dispersed λ between 0 and 1 for its starting values in the 3 chains. Values for Σ_b, σ^2_y were chosen such that starting values were far above and below the final region of the posterior distribution.

3.5 Results

We generated 3 chains of length 500,000, with the first 50,000 draws as burnin, and thinned every 15 iterations to save on memory and reduce the autocorrelation of parameter draws. Traceplots and Gelman-Rubin diagnostics (all < 1.05) indicated good mixing and no signs of non-convergence. The Monte Carlo standard error was calculated using the R package `mcmcse`. The MC error was less than 1.5% of posterior standard deviation for all parameters.

3.5.1 Parameter estimates

Figure 3.6 shows posterior means and 95% credible intervals for all regression coefficients for the model for Y_1 and Y_2. The signs on the coefficients and statistical significance levels nearly match for all covariates across the two parts of the model. Males and those with physical jobs tend to exhibit a higher number of bouts per day and more average excess MET-minutes per bout. BMI is negatively associated with bouts per day as well as average excess MET-minutes per bout. Age is negatively associated with bouts but not with average excess MET-minutes per bout. Having a college education was positively associated with average excess MET-minutes per bout, but not with number of bouts. Being Hispanic did not seem to have a major effect on either mean function. Being Black, however, is negatively associated with average excess MET-minutes but has no effect on number of bouts. Table 3.3 shows posterior means and 95% credible intervals for the remaining parameters. Recall that in a Generalized Poisson distribution, a value of $\lambda > 0$ indicates overdispersion. For the PAMS data, the estimate of λ was 0.220 (0.19,0.25) indicating that overdispersion is present. The estimated measurement error variances, $\sigma^2_{b_1}, \sigma^2_{b_2}$, are large relative to the regression coefficients corresponding to their respective model component. This suggests that there is considerable day to day variation in physical activity and that the device measurements themselves are noisy. The estimate of ρ_b is 0.411 (0.279,0.535), indicating that there is a significant amount of correlation between the mean functions of Y_1 and Y_2.

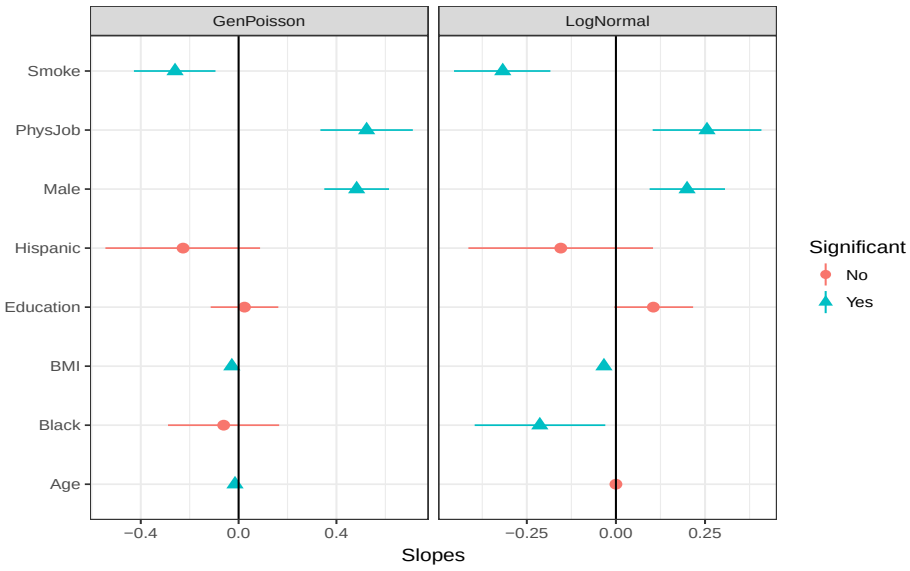

Figure 3.6: Posterior means and 95% credible intervals for regression coefficients for both parts of model.

Table 3.3: Posterior mean and 95% credible intervals for variance and overdisperision parameters

	Mean	Lower 95%	Upper 95%
λ	0.222	0.190	0.255
σ_y^2	0.466	0.418	0.520
$\sigma_{b_1}^2$	0.692	0.601	0.792
$\sigma_{b_2}^2$	0.274	0.214	0.337
ρ_b	0.411	0.279	0.535

3.5.2 Distribution of usual MVPA

In a previous section we explained how we can generate distributions of MVPA in MET-minutes for any population of interest. For illustration, we consider three subpopulations, and obtain the usual distribution of MVPA per day, for the average person in each group. Individual 1 is a 30 year old white female, BMI of 24, non-smoker, college graduate with a job that is not physically demanding. Individual 2 is a 65 year old black male, BMI of 35, smoker, non-college graduate also with a non-physical job. Individual 3 is a 40 year old white male, BMI of 29, smoker, no college degree and a physically demanding job. Figure 3.7 shows the distribution of usual MVPA

in MET-minutes for the population of individuals who have the same characteristics as those listed above. Pointwise 95% credible intervals are given to show uncertainty in estimation. The vertical dotted line represents the PAG recommendation for activity. The mean compliance rates and 95% credible intervals for individuals with these characteristics are 0.901 (0.831, 0.961), 0.413 (0.213, 0.640), and 0.967 (0.927, 0.993), respectively. Figure 3.8 shows the distribution of MVPA in MET-minutes for the PAMS sample on top, and for the four county Iowa population on bottom. We calculated quantiles for the population in the four Iowa counties surveyed using the survey weights with the R function `wtd.quantile` in the `Hmisc` package. The mean compliance rates and 95% credible intervals for the PAMS sample was 0.745 (0.710, 0.780), and the for the entire population of the four counties surveyed, the estimate was 0.787 (0.746, 0.826). Table 3.4 gives quantiles and standard deviations for the distribution of MET-minutes in MVPA for the PAMS sample, four county Iowa population, and the three average individuals described above.

Table 3.4: Estimated distribution of MET-minutes in MVPA for PAMS sample, Population of four Iowa counties, and three individuals. Estimated percentiles are means, and below its estimate is the standard deviation.

	5th	10th	25th	50th	75th	90th	95th	PAG Comply
PAMS Sample	17.76	27.93	63.43	165.48	422.60	940.45	1498.47	0.745
	2.07	2.79	5.21	11.17	29.75	81.68	159.47	(0.710,0.782)
IA 4 County	21.10	33.50	76.80	198.16	493.54	1083.37	1728.28	0.787
Population	2.73	3.75	7.38	16.55	43.08	115.96	237.39	(0.746,0.826)
Individual 1	52.45	83.43	171.13	377.41	815.53	1555.39	2336.90	0.901
	13.81	21.77	43.24	90.85	189.95	354.18	534.12	(0.831,0.961)
Individual 2	27.21	42.12	84.94	187.11	406.72	784.38	1177.38	0.413
	8.51	13.84	28.49	61.94	129.13	243.07	363.75	(0.213,0.640)
Individual 3	87.89	136.18	266.39	555.06	1152.71	2169.42	3252.25	0.967
	24.05	36.18	68.48	135.61	281.79	532.56	803.84	(0.927,0.993)

3.5.3 Usual MVPA using NHANES data

The National Health and Nutrition Examination Survey is a large national survey that can be used to assess the health of Americans. From 2003-2006 NHANES included physical activity monitoring with accelerometers (ActiGraph AM-7164). We wished to compare the results we

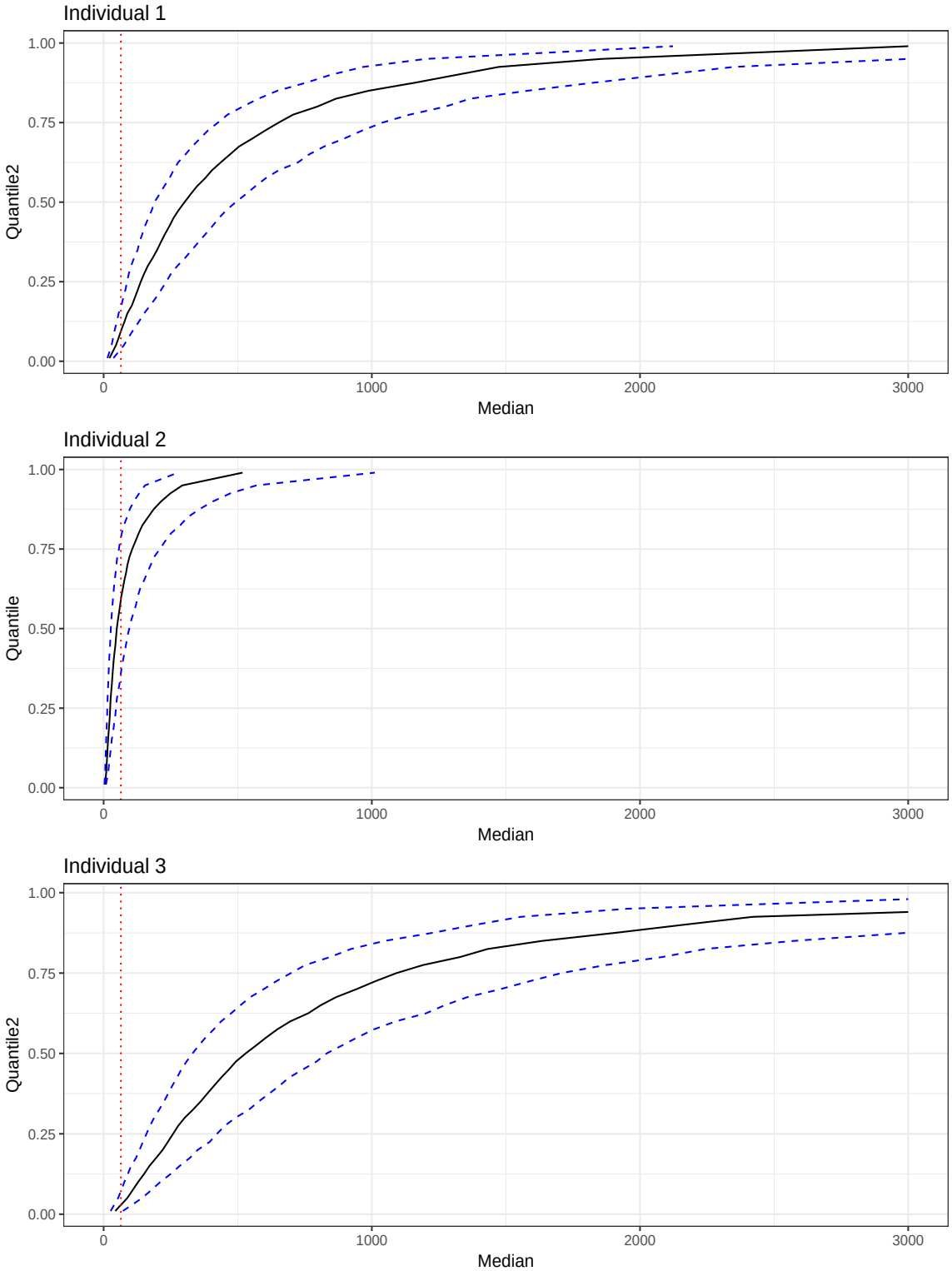

Figure 3.7: CDFs of of usual MET-minutes for 3 individuals with specified characteristics. Pointwise 95% credible intervals are given to show uncertainty in parameter estimates.

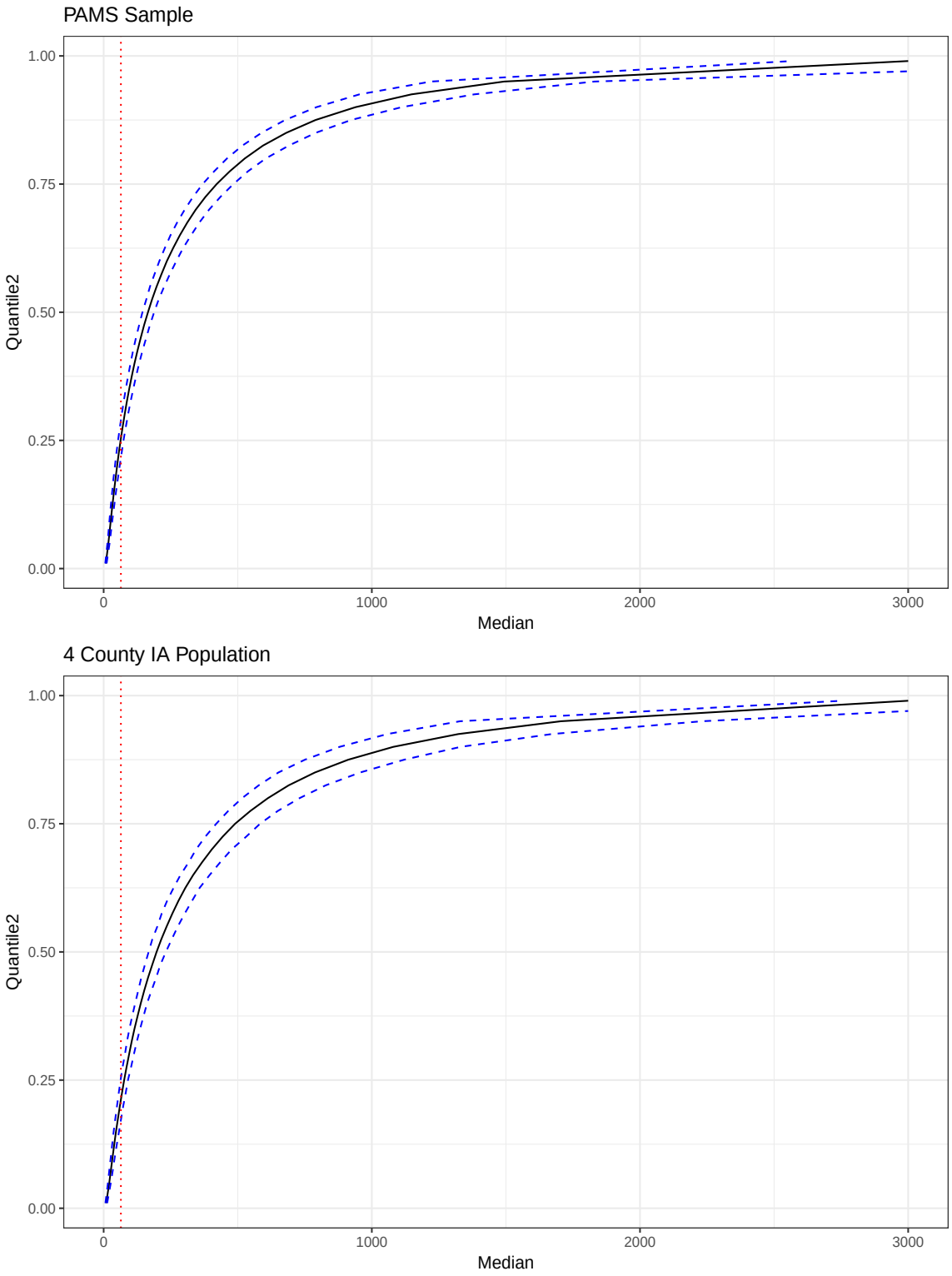

Figure 3.8: CDFs of of usual MET-minutes for PAMS sample and for population of four Iowa counties. Pointwise 95% credible intervals are given to show uncertainty in parameter estimates.

obtained from the PAMS study to a different large survey that collected accelerometery information. So that results obtained from the two surveys would be comparable, we used the method proposed by Ho et al. (2007) to select a subsample from the NHANES participants of equal size to PAMS and that would match the PAMS sample in other important ways. We implemented the method using their R package `MatchIt`. The subsample from NHANES was selected such that each person in PAMS was matched to someone from NHANES on demographic variables including gender, age, race, education, and BMI. Unfortunately, NHANES does not report participants' occupation, a variable that we found to be significantly associated with physical activity. For the individuals we include from NHANES, we randomly sampled two days of accelerometer measurements from the six available days. To compute bouts for the NHANES participants, we used the minute to minute information and follow the approach suggested by Tucker et al. (2016), and the threshold for moderate activity to be 2020 counts per minute . Counts during minutes within bouts were then converted to MET-minutes using the method of Freedson et al. (1998). We fit the model we propose to the NHANES data. Estimated distributions of MET-minutes in MVPA for the US population as well as for the three average individuals based on NHANES are shown in Table 3.5. The results for NHANES are similar to those in Tucker et al. (2011), but there is a large difference between the results in Tables 3.4 and 3.5. Levels of activity appear to be much lower when the model is fit to NHANES data. These large differences may be attributed o several differences between PAMS and NHANES: i) PAMS is a sample of the the population of four Iowa counties while NHANES is a nationally representative sample, ii) PAMS used the SWA to measure physical activity while NHANES used the Actigraph accelerometer (these instruments are designed to capture different indicators of physical activity), iii) compliance and wear time were much higher for PAMS, iv) the SWA uses a proprietary algorithm to calculate METs while we used Freedson et al.'s method to compute METs for NHANES. Finally, over 10 years elapsed between the two surveys. Consequently, we can expect differences in terms of the desirability of participating in physical activity. Although the populations from which the samples were drawn are not comparable, we would not expect such a large difference between the two populations. Participants in PAMS wore their monitor for

the entire day and night while NHANES participants were instructed to wear the device during waking hours, so this difference in wear time should not have a major effect on the measurement of MVPA. We believe that the major differences can be at least partially attributed to the variability in different brands of accelerometers and they way in which they convert movement to activity levels/METs. Crouter et al. (2006) gives a summary of the many methods proposed to convert counts to METs, showing there is a large variety of methods and considerable variation between the methods.

Table 3.5: Estimated distribution of MET-minutes in MVPA for NHANES subsample comparable to PAMS, US Population, and three individuals. Estimated percentiles are means, and below its estimate is the standard deviation.

	5th	10th	25th	50th	75th	90th	95th	PAG Comply
NHANES Sample	0.06	0.15	0.70	3.95	24.55	166.84	679.39	0.158
	0.03	0.06	0.21	0.84	4.35	39.89	251.99	(0.129,0.188)
US Population	0.07	0.16	0.74	4.18	26.20	181.26	758.97	0.163
	0.03	0.06	0.22	0.91	5.02	49.87	318.93	(0.132,0.195)
Individual 1	0.15	0.29	1.40	8.16	66.58	345.43	940.97	0.248
	0.04	0.07	0.37	2.54	23.30	101.32	242.85	(0.182,0.323)
Individual 2	0.01	0.02	0.09	0.41	1.98	8.46	27.48	0.030
	0.00	0.01	0.04	0.18	0.95	4.68	16.83	(0.006,0.056)
Individual 3	0.04	0.07	0.31	1.46	8.21	40.35	136.43	0.076
	0.01	0.02	0.09	0.46	3.11	16.72	55.36	(0.047,0.119)

3.6 Model Assessment

To assess how well our model fits the data, we generated $M = 1000$ replicate data sets from the posterior predictive distribution and compared selected statistics computed from the replicated datasets and from the original sample. From these comparisons, we can calculate posterior predictive p-values and other model diagnostics.

The posterior predictive distribution for the response vector $\mathbf{Y}$ is given by:

$$
\begin{aligned}
p(\mathbf{Y}^*|\mathbf{Y},\mathbf{Z}) &= \int\int\int p(\mathbf{Y}^*|\mathbf{b_1},\mathbf{b_2},\mathbf{Z},\boldsymbol{\theta_1})p(\mathbf{b_1},\mathbf{b_2}|\mathbf{Z},\boldsymbol{\theta_2})p(\boldsymbol{\theta}|\mathbf{Y},\mathbf{Z})d\boldsymbol{\theta}d\mathbf{b_1}d\mathbf{b_2} \\
&= \int\int\int p(\mathbf{Y_1^*},\mathbf{Y_2^*}|\mathbf{b_1},\mathbf{b_2},\mathbf{Z},\boldsymbol{\theta_1})p(\mathbf{b_1},\mathbf{b_2}|\mathbf{Z},\boldsymbol{\theta_2})p(\boldsymbol{\theta}|\mathbf{Y_1},\mathbf{Y_2},\mathbf{Z})d\boldsymbol{\theta}d\mathbf{b_1}d\mathbf{b_2} \\
&= \int\int\int p(\mathbf{Y_1^*}|\mathbf{b_1},\mathbf{Z},\boldsymbol{\theta_1})p(\mathbf{Y_2^*}|\mathbf{b_2},\mathbf{Z},\boldsymbol{\theta_1})p(\mathbf{b_1},\mathbf{b_2}|\mathbf{Z},\boldsymbol{\theta_2})p(\boldsymbol{\theta}|\mathbf{Y_1},\mathbf{Y_2},\mathbf{Z})d\boldsymbol{\theta}d\mathbf{b_1}d\mathbf{b_2},
\end{aligned}
\tag{3.13}
$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$ represents all unknown parameters except the latent variables. To simulate a new data set, we generate M random draws from the posterior distribution $p(\boldsymbol{\theta}|\mathbf{Y_1},\mathbf{Y_2},\mathbf{Z})$. We use the draw as well as the values of the covariates to simulate new latent variables $\mathbf{b_1},\mathbf{b_2}$ from $p(\mathbf{b_1},\mathbf{b_2}|\mathbf{Z},\boldsymbol{\theta_2})$. Finally, with the latent variables and posterior draws, we simulate new observations $\mathbf{Y_1}$ from the data model $p(\mathbf{Y_1^*}|\mathbf{b_1},\mathbf{Z},\boldsymbol{\theta_1})$ and $\mathbf{Y_2}$ from $p(\mathbf{Y_2^*}|\mathbf{b_1},\mathbf{b_2},\mathbf{Z},\boldsymbol{\theta_1})$. A posterior predictive p-value for statistic $T(Y,\boldsymbol{\theta})$ is calculated by:

$$
p - value = \frac{1}{M}\sum_{i=m}^{M} I(T(Y_m^*,\boldsymbol{\theta}_m) < T(Y^{obs},\boldsymbol{\theta})). \tag{3.14}
$$

To assess the fit of Y_1, we count the number of individuals who had zero bouts on day one and zero bouts on day two, the number of individuals who had one bout on day one and zero bouts on day two, and so on for all combinations of 0,1,2+ bouts. We stop at 2+ because if an individual has two bouts in a day, they will almost certainly achieve the recommended time in MVPA. Doing this for all $M = 1000$ simulated data sets, we calculate means for each category across all simulated data sets and compare to our observed proportions using a Chi-square test for proportions. Table 3.6 shows the results. The large p-value here indicates that data simulated from the fitted model look similar to the observed data, at least with respect to the specific statistic.

We also calculate the mean within-person standard deviation of Y_1 and the within-person range of Y_1. Figure 3.9 shows the results of these tests where the observed value is indicated by the vertical line. The posterior predictive p-values for these are 0.7 and 0.139, respectively, which indicates no lack of fit. This along with the significant value for the overdispersion parameter λ justifies our choice of the Generalized Poisson over the regular Poisson distribution.

Table 3.6: Chi-square test for proportions comparing the mean values from simulated data sets to observed values for Y_1. $\chi^2 = 3.7705$, df=8, p-value = 0.8772

No.	bouts	Observed	Posterior Mean
0	0	126	132
1	0	57	65
2+	0	77	78
0	1	65	65
0	2+	71	78
1	1	48	46
1	2+	81	83
2+	1	91	83
2+	2+	441	424

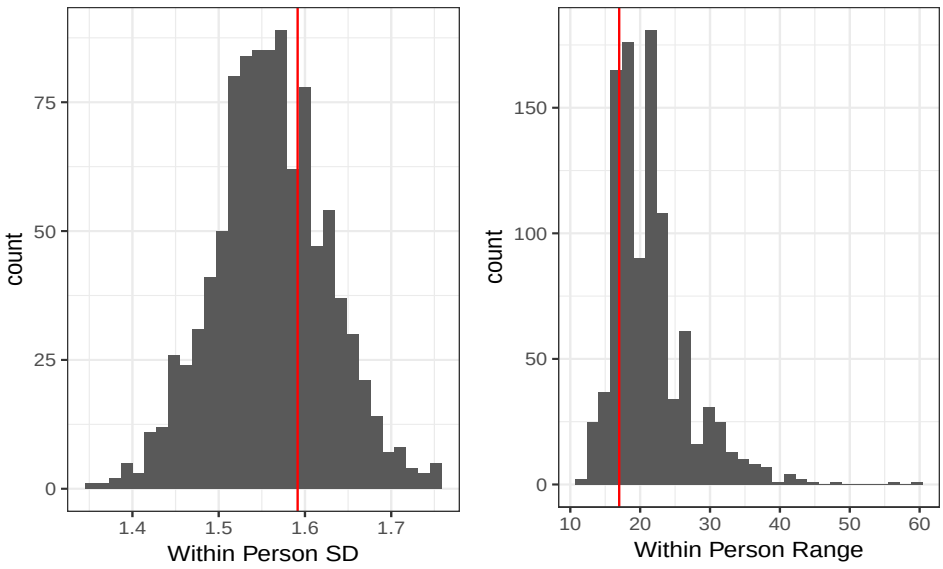

Figure 3.9: Posterior predictive model assessments. On left, within person standard deviation for Y_1. On right, within person range for Y_1. Red vertical line indicates truth.

Table 3.7: Summary of $M = 1000$ Kolmogorov-Smirnov test p-values comparing simulated data sets from fitted model to the observed data.

Q1	Q2	Q3	Mean
0.129	0.377	0.693	0.417

To assess the overall fit of the non-zero values of Y_2, we use the Kolmogorov-Smirnov test to compare each simulated data sets' empirical cumulative distribution function (ecdf) from the fitted model to the observed values' ecdf of Y_2. We perform this test for all M simulated data sets, so we have M p-values. Table 3.7 shows a summary of those p-values. These results show there are not apparent issues in the fit of Y_2 either. We performed the same model assessment procedures after fitting the model to the NHANES data, and the results were similar, indicating that the model also appears to fit the NHANES data well.

3.7 Discussion

In this paper we presented a two-part hierarchical model with measurement error that can be used to estimate distributions of usual MET-minutes in MVPA. In turn, the model can further be used to estimate compliance with the PAG. All parameter estimates were obtained using a Bayesian approach. We were able to accommodate the recommendation that activity must come in at least 10 minute bouts by jointly modeling the number of bouts and average excess MET-minutes per bout for individuals. Additionally, we modeled these as functions of demographic variables which could then be used to create distributions for subpopulations. We used data from the PAMS study to fit the model. In PAMS, participants wore an activity monitor on two separate days, for 24 hours. In preliminary analyses, we found that the 2-3 week buffer between measurements in PAMS seemed to successfully remove any dependence between recording days. The results showed that men and those with jobs that are physically demanding had higher levels of MVPA, and those with college degrees did as well but to a lesser extent. Age, BMI and being Black were negatively associated with MVPA. This type of information might be useful in designing interventions and that target specfic subpopulations.

The estimated distributions of usual MVPA that were based on the PAMS data were unexpected in that about 78% of the Iowa adult population meet the current physical activity guidelines. The high proportion of compliers is at odds with the rates of obesity and the sedentary lifestyle that have been documented (Owen et al. (2010)). Note that looking at Figure 3.4, based on the raw data, only 27% of the sample didn't achieve sufficient condition of two bouts per day to meet PAG guidelines. Moreover, only 12% didn't participate in a bout of MVPA. There are various interpretations for these results. First, obesity and physical activity may not be as tightly related as we tend to think and instead have a more complex association. It is possible that the PAG are set at a level that is easy to meet and that health benefits are realized with higher levels of physical activity. Another possible explanation is that the problem is more complicated, and that the relationship between obesity and physical activity, is mediated by diet, lifestyle, social, and genetic factors. To the best of our knowledge, no available dataset includes information we would need to capture such a complex relationship. In contrast, the results we obtained using the NHANES data suggest that only 16% of American adults are in compliance with the PAG. Additionally, the SWA is known to overestimate MVPA (Scheers et al. (2013), Santos-Lozano et al. (2017)). New methods that do not assume unbiasedness of accelerometry measurements are needed. To fit these new models, we require a gold standard to measure minute by minute physical activity in order to calibrate accelerometry measurements. To further complicate things, Hills et al. (2014) claims that it is "unlikely that a single measure of reported PA would suffice", in reference to assessing every possible activity humans engage in. Finally, recall that the PAG also advises adults to participate in two sessions of muscle buidling activity per week to realize health effects. We did not specifically measure this type of physical activity in PAMS, and thus did not consider it in our calculation of compliance rates. This may partially explain the high compliance to PAG, yet high obesity rates.

Our model can be fitted to different data sets, including national surveys like NHANES. The model permits incorporating sampling weights and the mean functions are flexible. To fit the model we propose, one must have replicate observations on at least a sample of individuals that can reasonably be assumed to capture daily physical activity unbiasedly. Further, it is important

to check whether the replicate measurements on an individual can reasonably be assumed to be exchangeable; if not, we need to transform the observations or extend the model to account for the correlation between daily measurements within a person. What differentiates our model from previous attempts to assess compliance to PAG is: i) our model accounts for measurement error in physical activity measurements, ii) it quantifies MVPA in MET-minutes rather than "minutes in MVPA", iii) it takes into consideration the 10 minute bout requirement, and iv) estimation is fully Bayesian which entails computation of easy to interpret posterior distributions for any quantity of interest.

CHAPTER 4. THE RELATIONSHIP BETWEEN MODERATE TO VIGOROUS PHYSICAL ACTIVITY AND METABOLIC SYNDROME: A BAYESIAN MEASUREMENT ERROR APPROACH

4.1 Introduction

Epidemiologists and health care professionals, among others, are interested in the relationship between physical activity and metabolic syndrome. Metabolic syndrome (MetS) is a group of risk factors that increase the chance of cardiovascular disease and type 2 diabetes (Kim et al. (2011)). The risk factors that comprise MetS include abdominal obesity (waist circumference >102 cm in men and >88 cm in women), high triglyceride level (>150 mg/dL), high fasting glucose level (>110 mg/dL), high systolic (>130 mm Hg) and/or diastolic blood pressure (>85 mm Hg), and high (low) levels of LDL (HDL) cholesterol (> 160 mg/dL and <40 mg/dL respectively) (Sisson et al. (2010), Tucker et al. (2016)). A person is said to have MetS when he or she exhibits three or more elevated risk factors as described above. MetS is a serious problem in the United States as it affects 34% of US adults (Moore et al. (2017)).

There have been numerous studies examining the relationship between physical activity and MetS. Franks et al. (2004) used multiple regression and logistic regression modeling and found a moderate negative association between level of physical fitness and MetS. Ford et al. (2005) and Dalacorte et al. (2009) both calculated odds ratios for MetS adjusted for covariates such as age and sex for prespecified levels of physical activity. They considered different population subgroups and reached different conclusions. Ford et al. (2005) focused on U.S. adults and reached similar conclusions as did researchers mentioned earlier. Dalacorte et al. (2009) considered Brazilian elders and found no associations between physical activity levels and MetS. Sisson et al. (2010) explored the relationship between steps walked in a day and odds of having MetS, and found that the odds of having MetS were 10% lower for each additional 1000 steps. Camhi et al. (2011) looked at the

relationship between moderate intensity lifestyle activity (lower than moderate to vigorous physical activity) and MetS and found that more time spent in this light activity, independent of MVPA, is associated with lower odds of MetS. Tucker et al. (2016) examined the relationship between self-reported and accelerometer-measured physical activity and MetS. Using logistic regression, they found that individuals who do not meet the *Physical Activity Guidelines for Americans* have greater odds of having MetS, (odds ratio of 2.57 for men and 4.40 for women when using accelerometer data). Huang et al. (2017) examined the odds ratios of MetS and low/high levels of leisure, occupational, and commuting physical activity in Taiwanese workers and found leisure and occupational physical activity to be associated with MetS. Tucker et al. (2016) and Huang et al. (2017) were the only papers that looked at odds ratios of individual MetS risk factors as well as MetS. Franks et al. (2004) was the only paper that provided an adjustment for physical activity due to measurement error.

Most of these studies considered the overall presence of MetS and physical activity levels for an individual, typically using cut-points for what is considered "high levels" of risk factors and using a criterion similar to the "three or more" to classify someone as having MetS. Then they use logistic regression using physical activity levels from accelerometers or self-reports to calculate odds ratios for different subpopulations. We propose to model these risk factors on the continuous scale as they are collected, rather than on the typical "high" or "normal" levels they are often attributed. Additionally, we investigate the relationship between physical activity and all the risk factors mentioned earlier. In this way, we can allow for non-linear and different relationships with MVPA for each factor. We also incorporate dependence in MetS risk factor measurements within an individual via seemingly unrelated regressions, since these measurements are not independent, even though all of the analyses we found in the literature assumed independence. Unlike most of the studies we have identified, we construct a measurement error model for the physical activity observations before we assess their relationship with MetS risk factors. Finally, our model provides a way to account for sample survey weights. Like many previous studies, we also look at the relationship between minutes of MVPA and MetS risk factors. Our goal is to understand the

Table 4.1: Demographic summaries for the NHANES sample.

	Men (n=4038)	Women (n=4206)
Age n		
18-34 years	1188	1405
35-49 years	958	972
50-65 years	918	954
66-85 years	974	875
Race/Ethnicity n		
Mexican American	846	909
Other Hispanic	107	120
Non-Hispanic White	2030	2034
Non-Hispanic Black	927	948
Other Race	128	195

complex relationships between MVPA and MetS risk factors while allowing for complex relationships that exist between the risk factors themselves.

4.2 Data

4.2.1 National Health and Nutrition Examination Survey

We use data from the National Health and Nutrition Examination Survey (NHANES) from 2003-2006. NHANES is a large, national survey that is used to assess the health of adults and children in the United States. As part of this health assessment, NHANES gathers information about the components of MetS described in the introduction for each participant, and also collects physical activity information using accelerometers. For our analysis, we focus on adults age 18+. Table 4.1 shows summary statistics for NHANES participants who are 18 years of age or older.

NHANES (CDC (2003-2006)) participants were instructed to wear the ActiGraph AM-7164 for seven straight days. Participants were instructed to wear the device during waking hours, and to remove the device only while sleeping or during any water activities. The NHANES website states that the device records uniaxial movements, which means that activities including stationary bikes, ellipticals, or primarily upper body movements may not be recorded accurately. The accelerometer records measurements of the intensity of activity minute by minute. The intensity of the activity

Table 4.2: MetS risk factor mean values (sd) for the NHANES sample.

	Men	Women
Waist Circumference (cm)	99.04 (14.96)	94.65 (15.44)
Glucose (mg/dL)	106.54 (32.76)	102.52 (31.80)
Triglycerides (mg/dL)	147.16 (115.15)	129.21 (83.51)
Systolic Blood Pressure (mm Hg)	125.23 (16.68)	122.07 (21.15)
Diastolic Blood Pressure (mm Hg)	69.52 (13.73)	67.35 (13.31)
LDL (mg/dL)	115.19 (35.11)	113.49 (37.34)
HDL (mg/dL)	49.84 (13.91)	60.97 (16.37)

is indicated by "counts" where more counts correspond to more intense activity, and zero counts correspond to no activity. If the number of counts in a minute exceeds 2020, then the person is engaged in moderate physical activity. If the number of counts exceeds 5999, the activity is said to be vigorous (Tucker et al. (2016)). We calculate daily minutes in MVPA for each individual using these thresholds. The resulting data set contains 7.8% instances of zero minutes.

Many of the NHANES participants who wore the accelerometer had some or all of the MetS risk factors measured as well. Most individuals have data on waist circumference, HDL, and systolic and diastolic blood pressure, but less than half (42%) have data on glucose, triglycerides, and LDL. We removed 46 individuals with a measured diastolic blood pressure of 0 from our analysis. Table 4.2 gives summaries of MetS risk factors by gender.

4.2.2 Adjustment for day worn and weekend effects

We use the values 1-7 to identify each of the seven days during which individuals wore the accelerometer. Let W_{0ij} represent the minutes in MVPA for individual i on day worn j. Table 4.3 shows the average minutes in MVPA for each of the days worn. The first five days were similar in terms of average minutes in MVPA. The sixth day drops a little and the seventh day has an even sharper drop. Although we can speculate why this is happening, discussions with other users of NHANES produced no definitive explanation. Additionally, we did not find any mention of this in papers that used the NHANES accelerometry data. Because of this, we decided to remove day 7 measurements from our analysis. We include day worn 6 in the analysis but adjust those

measurements so that they more closely resemble the rest. We also found a significant weekend effect on minutes in MVPA. On average, participants did 23.28 minutes of MVPA on weekdays (Monday-Friday) and 18.51 minutes of MVPA on weekends.

Table 4.3: Average minutes in MVPA and standard error across all participants by the number of days worn.

Day Worn	Mean (min)	Standard Error (min)
1	22.601	0.362
2	23.018	0.368
3	22.968	0.379
4	22.414	0.366
5	22.244	0.377
6	21.155	0.377
7	19.420	0.381

We consider the day worn and weekend effects to be nuisance effects. To adjust for them, we fit an additive model containing an indicator for day worn 6 versus all the others and an indicator for weekend versus weekday using OLS to observed minutes in MVPA, denoted W_{0ij} for individual i on day worn j. The adjustment is a ratio, rather than a linear adjustment and follows closely from Nusser et al. (1996). We let the ijth minutes in MVPA adjusted for day worn and weekend effects be $W_{1ij} = \hat{W}_{0ij}^{-1}\bar{W}_{0\cdot1}W_{0ij}$, where $\bar{W}_{0\cdot1}$ is the mean of the observed minutes in MVPA for days worn 1-5, and $\hat{W}_{0ij}$ is the predicted minutes in MVPA for individual i on day worn j from the regression above. This ratio adjustment method reduces the number of adjusted minutes in MVPA that are negative. If negative adjusted values occur, they are set to zero.

4.2.3 Missing accelerometer data

The total number of persons aged 18 years or older in NHANES 2003-2006 with accelerometry data is 8197. There is a non-negligible number of missing accelerometer measurements in NHANES. Some individuals are missing entire days, whereas others are missing time periods during the day. Analyses in the past using NHANES accelerometer data typically follow some general rule, like a "valid" day is considered as having at least 10 hours of wear time (Sisson et al. (2010), Tucker et al.

(2016)). In addition, some analyses only consider those individuals who have at least four valid days (Tucker et al. (2016)).

We consider two approaches to analyzing these data: first we focus on the subset of individuals with six days of accelerometery data with at least 10 hours of wear time each day. There are 3336 persons in this subgroup. In a second analysis, we will impute missing data within days as well as missing days so we do not cull individuals due to missing information. There are 7872 individuals in this subgroup. There have been many methods proposed in the literature for handling missing physical activity data (Catellier et al. (2005), Kang et al. (2009), Liu et al. (2016)); we use the multiple imputation approach developed specifically for minute by minute accelerometer data by Lee and Gill (2016).

Lee and Gill's method imputes missing data at the minute level. This way, if an individual has large gaps in data during a day in which they have some wear time, the imputation method can use that information rather than imputing the entire day from scratch. Lee et al.'s method considers three types of minute by minute values: i) positive counts indicating physical activity, ii) zero counts for less than or equal to P minutes, indicating no physical activity, and iii) zero counts greater than P minutes indicating non wear time. We select P to be 20 minutes. The imputation model assumes a zero inflated Poisson log-normal mixture model that also allows for autocorrelation. Using the notation of Lee and Gill (2016), the model for "counts" is :

$$Y_i|\lambda_i \sim \begin{cases} 0 & \text{with probability } \pi_{i,t} \\ \text{Poisson}(\lambda_i) & \text{with probability } (1-\pi_{i,t}) \end{cases} \tag{4.1}$$

$$\lambda_i \sim LN_d(x_i^T\beta, \Sigma) \tag{4.2}$$

$$\text{where} \tag{4.3}$$

$$\text{logit}(\pi_{i,t}) = x_i^T\gamma + \delta_t log(Y_{i,t-1}+1) \tag{4.4}$$

$$Z_t = log(Y_t) - x^T\beta \tag{4.5}$$

$$Z = (Z_{t-K}, ..., Z_{t-1}, Z_{t+1}, ..., Z_{t+k}) \tag{4.6}$$

and Σ is a $d \times d$ covariance matrix with $d = 2K + 1$. Imputation is carried out by setting the count to zero if $\pi_{i,t} > u$ where $u \sim$Unif(0,1), otherwise draw a Poisson random variable with parameter $\hat{\lambda}_i \exp(\hat{e})$, with $\hat{\lambda} = exp(x^T \hat{\beta})$ and $\hat{e} = \hat{\Sigma}_{yz} \hat{\Sigma}_{zz}^{-1} Z, \Sigma_{yz} = cov(Z_t, Z), \Sigma_{zz} = var(Z, Z)$. The imputation thus uses both K lag and lead effects. Their model can be implemented using their R package `accelmissing`. We create five imputed data sets.

4.2.4 Missing demographic information

We are interested in including the effect of demographic covariates in the model for MetS. The demographic variables we include as covariates are age, BMI, gender, race, and education level. There are no missing values for age, BMI, gender, or race. However, 274 individuals were missing data on education level, and instead of removing them we opt to impute education levels using a random forest. Education level is reported on a scale of 1-5, where 1 is less than 9th grade, 2 is 9-11th grade, 3 is high school graduate/GED, 4 is some college, and 5 is a college graduate. We used the `randomForest` package in R, and used variables on age, marital status, number of occupants in household, income, and poverty income ratio (a ratio of family income to poverty threshold) to pick the best classification for education level. Of the 274 individuals with no education level data, 47 also had no information on income so we removed them from the analysis, leaving us with 227 individuals with imputed values for education.

4.2.5 Sample survey weights

NHANES data are collected with an unequal probability survey design, meaning that some population groups are over- or under-represented. The survey weights in NHANES account for unequal probability of selection as well as for nonresponse. Both the 2003-2004 and 2005-2006 NHANES data provide sample weights for each individual. The sample weight can be thought of the number of people that sample one individual is representing. NHANES gives details[1] on how to adjust survey weights when using data from multiple NHANES surveys, such as we are. Since we are using two sources of information from NHANES, there are two possible places survey weights

[1] https://wwwn.cdc.gov/nchs/data/series/sr02_161.pdf

could affect our model. First, we use accelerometery data for the measurement error portion of our model. We will be modeling these data as a function of demographic covariates that help make up the survey weights calculations, so we will already be controlling for the variables that are used in the weighting (Gelman (2007)). The other part of our model uses the MetS risk factor data, and to incorporate this unequal sample data into our model which uses equal-weight methods, we take the approach used in Nusser et al. (1996) to adjust the observations. Nusser et al. (1996) uses the empirical cumulative distribution function to create an equal weight sample from the raw data. Using the notation of Nusser et al. (1996), the empiricial cumulative distribution function is calculated as:

$$\hat{F}_Y(a) = \sum_{i=1}^{n} w_i I_{Y_i}(a), \tag{4.7}$$

where $I_{Y_i}(a)$ is the indicator function and w_i is the survey weight corresponding to individual i. The adjusted value $Y_i = \hat{F}_Y^{-1}\left[(1/n)(s_i - 0.5)\right]$, where s_i is the rank of Y_i^*, the raw data. We do this adjustment for each of the seven MetS risk factors. All MetS risk factors hereafter refer to these adjusted equal-weight values.

4.2.6 Minutes of MVPA

The measurement error model that we propose, assumes that the measurement error is independent of true minutes of MVPA. Additionally, we with to assume minutes of MVPA are approximately normally distributed. We achieve both goals using a fourth root transformation and denote $W_{1ij}^{1/4} = Wij$. Figure 4.1 shows a normal quantile plot for the individual means of the fourth root of minutes in MVPA. Except for the spike at zero minutes, the transformed data appear to be approximately normally distributed. Figure 4.2 shows the distribution of the observed mean minutes in MVPA for each individual on the left pane. On the right pane, we show the observed mean minutes of MVPA transformed by the fourth root against their standard deviation. After transforming the observations, meaning the measurement error does not depend on minutes of MVPA. The parabolic shape on the left in the right plot is an artifact of the data; the curve arises

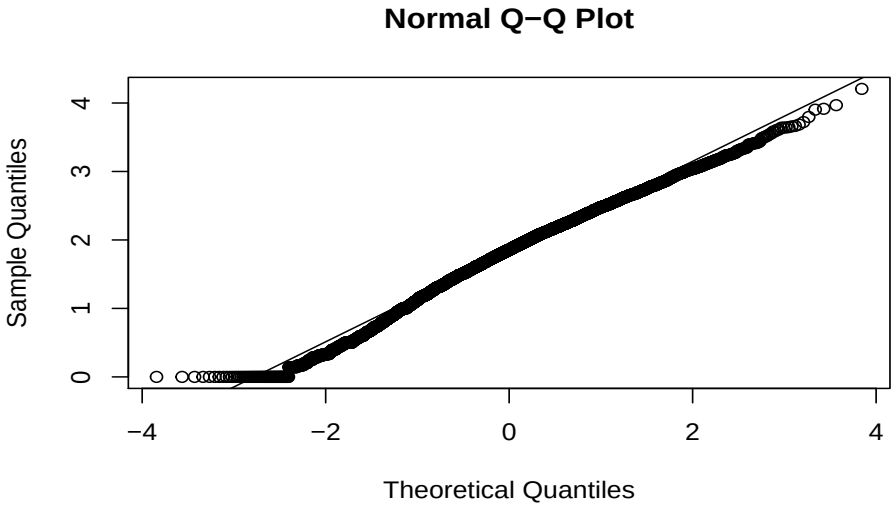

Figure 4.1: Normal quantile plot for individual means of fourth root of minutes in MVPA.

when the observed minutes in MVPA for an individual are either zero or one minute in each of the seven days. Based on this diagnostic, there does not appear to be dependence of the measurement error variance on activity levels.

To assess the relationship between minutes in MVPA and MetS risk factors, Figure 4.3 shows observed mean of fourth root of minutes in MVPA for each individual versus each risk factor. We plot a loess curve on top of each plot to get an understanding of the relationships. We log transformed Glucose and Triglycerides because the measurements are highly skewed.

4.3 Measurement Error and Regression Models

The main goal of this work is to understand the relationship between usual physical activity levels and MetS risk factors. We denote usual minutes in MVPA for individual i as t_i which is defined as:

$$t_i \equiv E(T_{ij}|i), \tag{4.8}$$

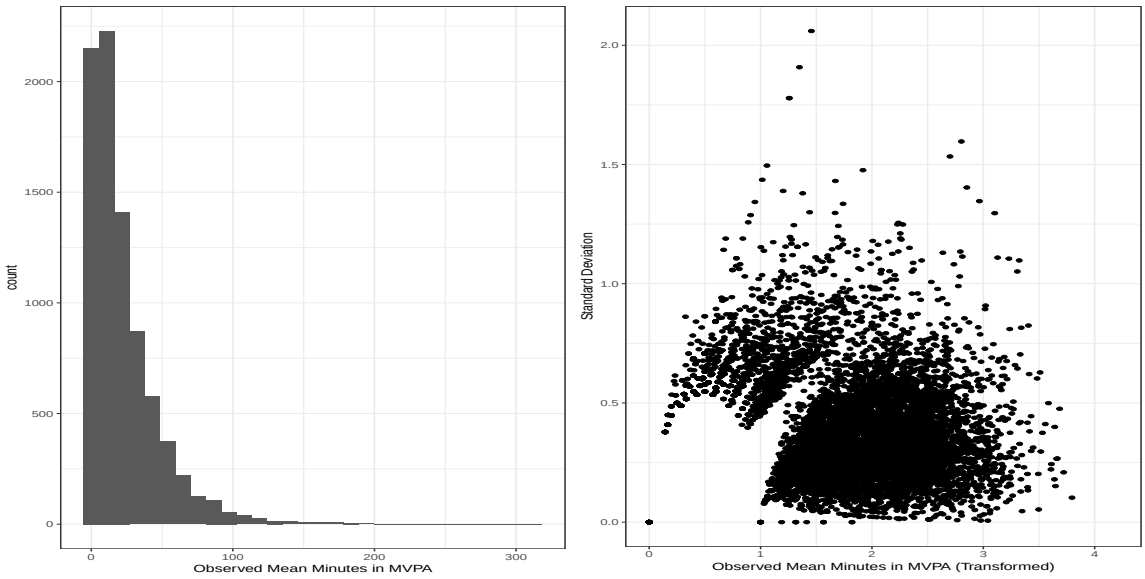

Figure 4.2: Left: Distribution of observed mean minutes in MVPA for each individual. Right: observed mean minutes of MVPA transformed by 4th root against standard deviation.

where T_{ij} is the true number of minutes in MVPA on day j for individual i. We denote the vector of MetS risk factors adjusted for sample weights for individual i by $\mathbf{Y_i}$, which has elements $Y_{\cdot,wst}$ denotes waist circumference in centimeters, $Y_{\cdot,glu}$ denotes log glucose (mg/dL), $Y_{\cdot,tri}$ denotes log triglycerides (mg/dL), $Y_{\cdot,ldl}$ denotes LDL cholesterol (mg/dL), $Y_{\cdot,hdl}$ denotes HDL cholesterol (mg/dL), $Y_{\cdot,bps}$ denotes systolic blood pressure (mm Hg), and $Y_{\cdot,bpd}$ denotes diastolic blood pressure (mm Hg). We model the relationship between t_i and Y_i through the regression model

$$E(\mathbf{Y_i}|t_i) = \mathbf{m}(\boldsymbol{\gamma}; t_i). \tag{4.9}$$

The difficulty is that we do not observe t_i or T_{ij}, rather we observe W_{1ij} as described in the previous section. This observation is contaminated with measurement error through error in the measurement device, day to day variability in physical activity, and possible nuisance effects like weekend or day worn (we already adjusted for these last two). First we present the measurement error model for observed minutes in MVPA that we propose, and then we present the regression model for the relationship between usual minutes in MVPA and MetS risk factors.

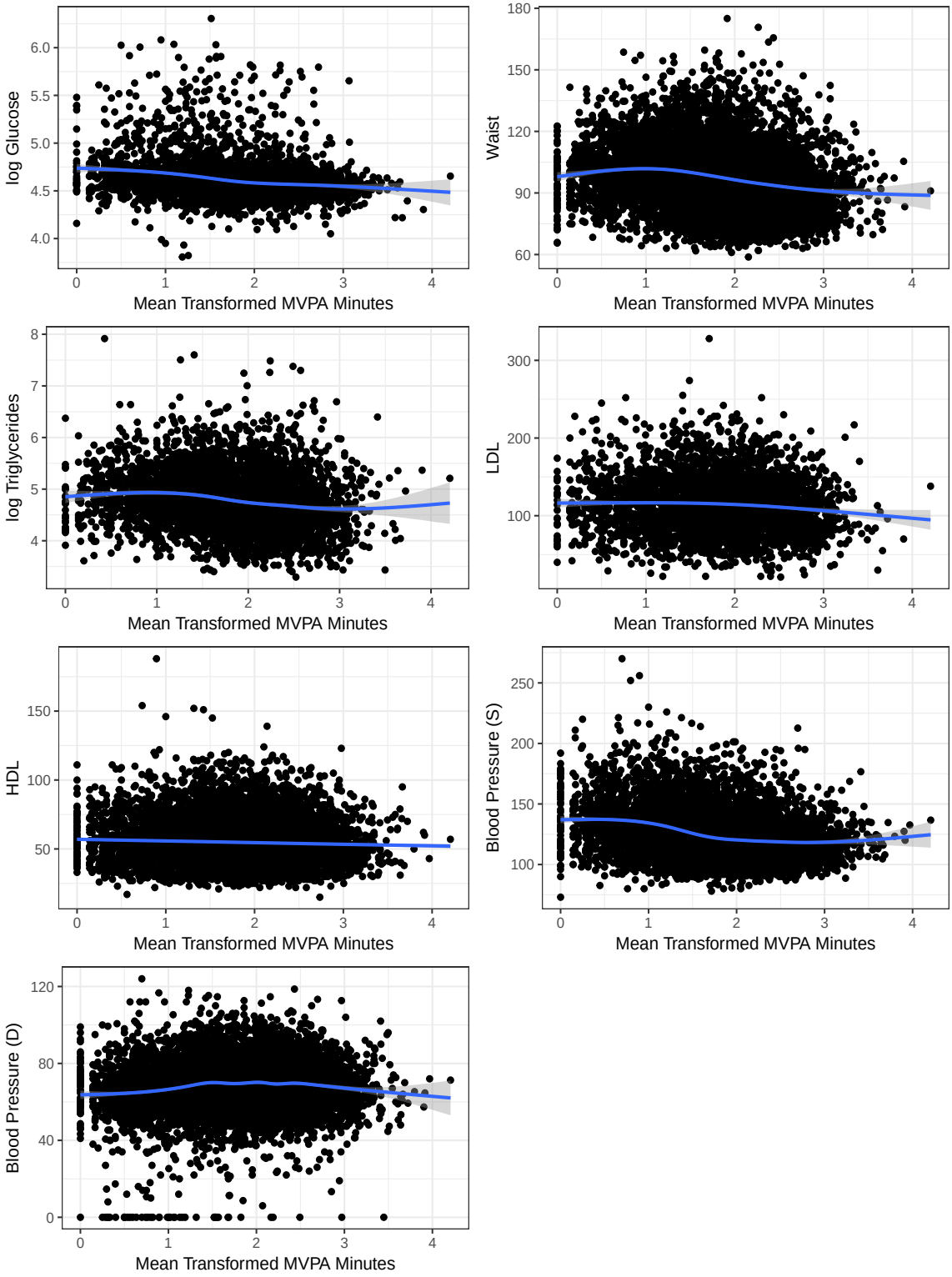

Figure 4.3: Observed mean transformed minutes in MVPA against MetS risk factors. Loess curve drawn on each plot to see trend.

4.3.1 Measurement error model

4.3.1.1 Model assumptions

The measurement error model we propose follows the work of Olsen and Schafer (2001), Tooze et al. (2006), and Kipnis et al. (2009) where observations take both positive and zero values. There are then two components in the model: the probability of an individual participating in MVPA, and given participation in MVPA, for how long. Individual i's usual minutes in MVPA can be further defined as:

$$t_i \equiv E(T_{ij}|i) = E(T_{ij} \times I(T_{ij} > 0)|i) = E(T_{ij}|T_{ij} > 0, i)P(T_{ij} > 0|i) = A_i\pi_i, \tag{4.10}$$

where A_i is the usual minutes of MVPA on a day where individual i participated in MVPA, and π_i is the probability that individual i participates in MVPA on any given day. Like Tooze et al. (2006) and Kipnis et al. (2009), we make the following assumptions: i) MVPA is reported by the accelerometer if and only if the individual actually participated in MVPA, ii) the accelerometer is unbiased for usual minutes of MVPA participation days. More formally:

$$E(W_{1ij}|i) = E(W_{1ij}|W_{1ij} > 0, i)P(W_{1ij} > 0|i) \overset{assume}{=} A_i\pi_i = t_i. \tag{4.11}$$

4.3.1.2 Model specification

As discussed previously, we transform our observed minutes of MVPA by taking a fourth root to remove dependence of measurement error on minutes of MVPA and to approximate normality. We model the π_i part with a logistic regression:

$$P(W_{1ij} > 0, |i) = \pi_i = logit^{-1}(\boldsymbol{\alpha}'\mathbf{Z_{1i}} + b_{1i}), \tag{4.12}$$

where b_{1i} is a person level random effect, and $\mathbf{Z_{1i}}$ is a vector containing covariate information for individual i. Our model for transformed, non-zero minutes in MVPA is of the form:

$$W_{1ij}^{1/4} = W_{ij} = \boldsymbol{\beta}'\mathbf{Z_{2i}} + b_{2i} + \epsilon_{ij}, \tag{4.13}$$

where b_{2i} is a person level random effect, assumed to be independent of ϵ_{ij}, and $\mathbf{Z_{2i}}$ is a vector containing covariate information for the second part of the measurement error model. In our work, we let $\mathbf{Z_{1i}} = \mathbf{Z_{2i}} = \mathbf{Z_i}$ be the covariate vector for individual i containing age, BMI, race, and education level. Because MVPA data from accelerometers from NHANES were collected across seven continuous days for each individual, assuming exchangeability for individuals' observations would not be appropriate. Instead, we allow for autocorrelation across the six days. Within an individual, we let the ϵ_{ij} have an AR(1) structure. We do not believe that there is this structure in the residuals of the model for π_i, since even if an individual is only active on some days, this is likely unrelated to the probability that they reach even one minute of MVPA on other days. Although we have already shown that the measurement error does not depend on minutes of MVPA, it is possible that it varies based on demographic variables. To check for this, we fit the mixed effects model in 4.13 with AR(1) correlation structure on ϵ_{ij} using maximum likelihood with the R function `lme`, and calculate residuals using the parameter estimates and best linear unbiased predictors of person random effects. We then calculated within group variances of the residuals split by gender, by race, by age, and by BMI. There were minor differences by gender, race, and BMI, but age had major differences. Figure 4.4 shows the estimated within group variance for persons aged between 18 and 85 years, with an overlaid loess fit to indicate the trend. It appears that individuals under 40 have fairly consistent MVPA levels, individuals between 40 and 60 have more variability in their MVPA levels, and individuals over 65 have increasingly less variation in their MVPA levels as they age. Based on this trend, we model the variance of the ϵ in 4.13 as a cubic function of age:

$$var(\epsilon_{ij}) = \sigma_i^2(\boldsymbol{\delta}) = \delta_0 + \delta_1 age_i + \delta_2 age_i^2 + \delta_3 age_i^3 + \nu_i. \tag{4.14}$$

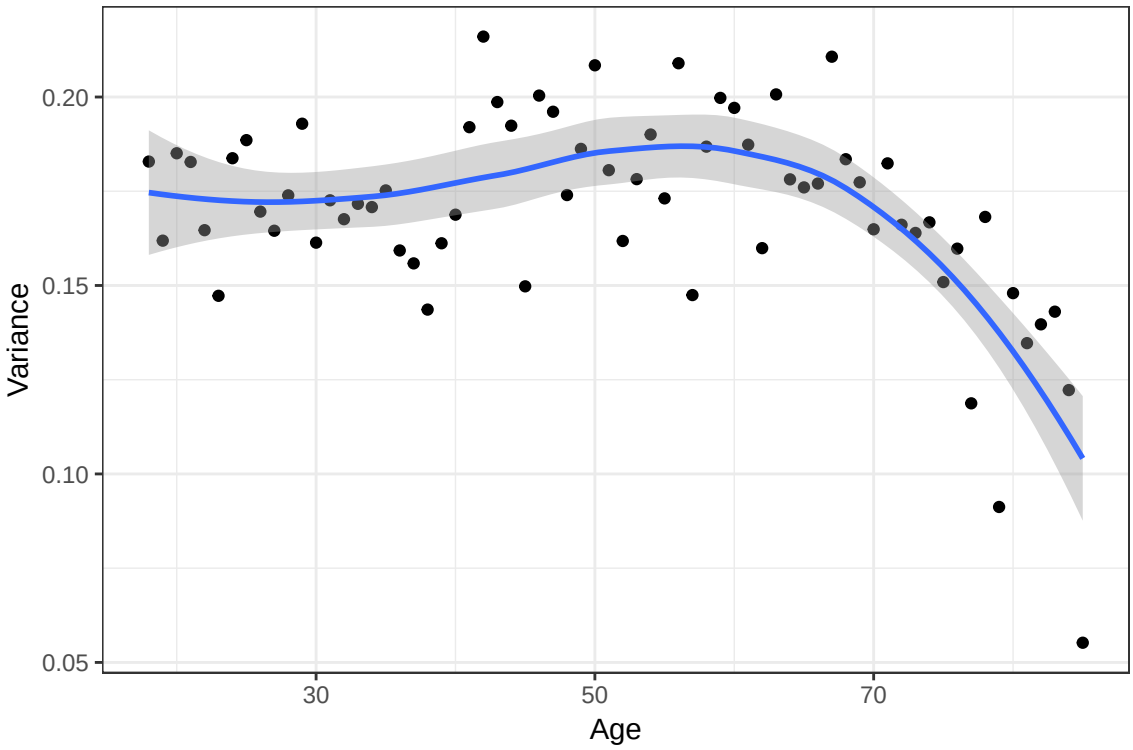

Figure 4.4: Variance by age for residuals of model in 4.13 with AR(1) correlation structure. Model was fit using only positive values of W_{ij}.

Table 4.4: OLS estimates for error variance parameters in 4.14.

	Estimate
δ_0	0.2662
δ_1	-.008
δ_2	2.09e-04
δ_3	-1.62e-06

We fit this model using ordinary least squares (OLS) and the coefficient estimates are in Table 4.4. We use these estimates as plug-in estimates as to not further complicate the MCMC and because we are not interested in inference for these parameters.

It is also possible that the AR(1) parameter in the error term is different for different demographic groups. We fit the model in 4.13 using maximum likelihood on different subgroups individually where for age we group individuals into 18-34, 35-49, 50-64, 65-85 age categories. Table 4.5 shows the estimates for the autoregressive parameter ϕ. Based on these results, we aggregate individuals into two categories: 18-64 and 65-85. We label group 1 as individuals under 65, and group 2 as individuals 65 and over.

Let $\mathbf{W_{i+}}$ and $\mathbf{C_i}$ be vectors of length m_i containing positive values and indices of day worn for positive values of W_{ij} for individual i, respectively. We then write our model as:

$$I(W_{ij} > 0)|\pi_i \overset{ind}{\sim} \text{Bernoulli}(\pi_i), \tag{4.15}$$

$$\pi_i|b_{1i}, Z_i = logit^{-1}(\boldsymbol{\alpha}'\mathbf{Z_i} + b_{1i}), \tag{4.16}$$

$$\mathbf{W_{i+}}|b_{2i}, Z_i \overset{ind}{\sim} N\left(\boldsymbol{\beta}'\mathbf{Z_i} + b_{2i}, \Sigma_i(\phi, \hat{\boldsymbol{\delta}})\right), \tag{4.17}$$

$$b_{1i}, b_{2i} \overset{ind}{\sim} N(\mathbf{0}, \Sigma_b), \tag{4.18}$$

where Σ_b is a 2×2 covariance matrix with diagonal elements $\sigma^2_{b1}, \sigma^2_{b2}$ and correlation ρ_b (note that this links (4.12) and (4.13)) ; $\Sigma_i(\phi, \hat{\boldsymbol{\delta}})$ is an $m_i \times m_i$ covariance matrix with elements $\sigma_{i,kl}$ in the kl^{th} position defined as:

Table 4.5: AR(1) parameter estimates from model in 4.13 for different groups based on demographic variables. Model was fit individually for each level of each grouping variable.

Group	Level	n	$\hat{\phi}$
Gender	Male	3924	0.128
	Female	4229	0.118
Race	Mexican-American	1735	0.161
	Other Hispanic	236	0.150
	White	4028	0.111
	African-American	1812	0.118
	Other Race	342	0.151
Age	18-34	2743	0.143
	35-49	1843	0.145
	50-64	1633	0.140
	65-85	1934	0.050
BMI	< 18	91	0.18
	18-25	2598	0.11
	25-30	2770	0.128
	30-35	1579	0.162
	35+	1115	0.115

$$\sigma_{i,kl} = \sigma_i^2(\hat{\boldsymbol{\delta}})\phi_{g(i)}^{|C_i[k]-C_i[l]|}, \tag{4.19}$$

where $C_i[k]$ denotes the k^{th} element of $\mathbf{C_i}$ and $\sigma_i^2(\hat{\boldsymbol{\delta}})$ is defined in 4.14.

4.3.1.3 Obtaining estimates of usual minutes in MVPA

To obtain an unbiased estimate of t_i, we use the relation in (4.11), but since we have transformed W_{1ij}, the expectation we wish to estimate is:

$$A_i = E(W_{1ij}|W_{1ij} > 0, i) = E(W_{ij}^4|W_{ij} > 0, i). \tag{4.20}$$

We approximate this expectation using Tayor's expansion:

$$E(f^{-1}(x)) \approx f^{-1}(x) + \frac{1}{2}\sigma_\epsilon^2 \frac{\partial^2(f^{-1}(x))}{\partial x^2}, \tag{4.21}$$

which in our setup is:

$$A_i = E(W_{ij}^4 | W_{ij} > 0, i) \approx (\boldsymbol{\beta}'\mathbf{Z_i} + b_{2i})^4 + 6\sigma_e^2(\boldsymbol{\beta}'\mathbf{Z_i} + b_{2i})^2. \tag{4.22}$$

Combining the assumptions in (4.11) and the results from (4.12) and (4.20,4.22), usual number of minutes in MVPA for individual i, t_i under our model, is given by:

$$t_i = A_i\pi_i \approx logit^{-1}(\boldsymbol{\alpha}'\mathbf{Z_i} + b_{1i}) \left((\boldsymbol{\beta}'\mathbf{Z_i} + b_{2i})^4 + 6\sigma_i^2(\hat{\boldsymbol{\delta}})(\boldsymbol{\beta}'\mathbf{Z_i} + b_{2i})^2\right). \tag{4.23}$$

4.3.2 MetS regression model

Once we have defined an estimator for t_i as in the measurement error model (4.23), the next step is to estimate the regression parameters of (4.9). This approach is similar to regression calibration (Caroll et al. (2006)), but there are some differences. We refer to our approach as *Bayesian regression calibration*. The major difference between regression calibration and Bayesian regression calibration is the following: for every draw from the full conditionals of the model parameters, we calculate t_i for every individual, and compute regression coefficients in 4.9 using these t values. Doing so automatically accounts for the variability in the regression coefficients in 4.23 used to calculate t_i. This is in contrast to classical regression calibration where one set of $\hat{t}$ values is computed using point estimates for parameters in (4.23), and then those $\hat{t}_i$ are plugged into (4.9). The latter approach requires using the bootstrap or other methods to adjust standard errors, whereas our Bayesian regression calibration does it all in one step.

We determine the form of the m functions to model the association between MVPA and each risk factor using the plots in Figure 4.3. Based on the form of the loess fit, waist circumference, log glucose, log triglyceride, and systolic blood pressure all appear to have a similar, upside down S shape. The function:

$$f(x) = M - \frac{L}{1 + e^{-K(x-B)}}, \tag{4.24}$$

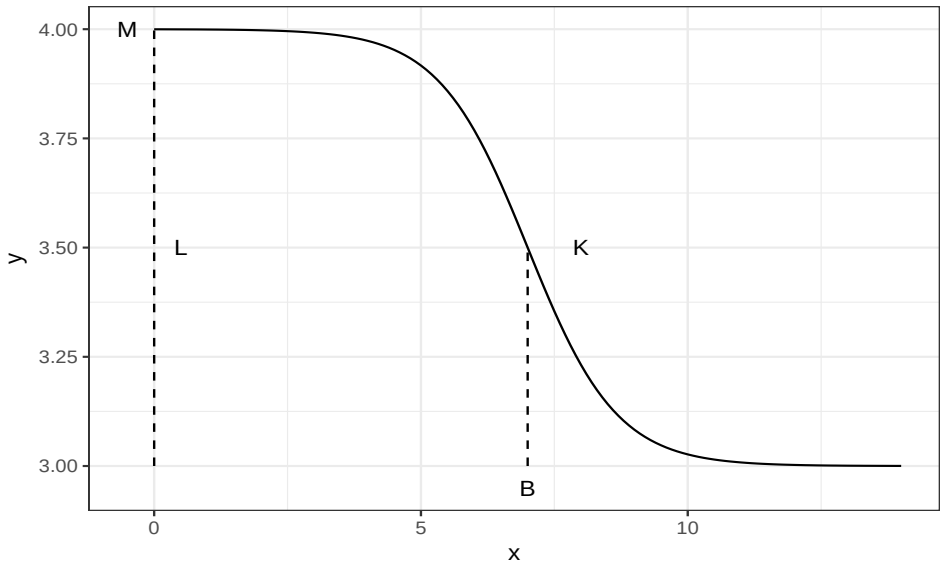

Figure 4.5: Nonlinear function for Waist circumference, Glucose, Triglyceride, and Systolic Blood Pressure.

has this form. We let $m_{wst}(), m_{glu}(), m_{tri}()$, and $m_{bps}()$ denote the functions with this form. Figure 4.5 gives an example of (4.24) with parameter values: $M = 4, L = 1, B = 7, K = 1.2$. One of the benefits of this functional form is that the parameters are interpretable: M is the limit of $f(x)$ as x goes to $-\infty$, $M - L$ is the limit of $f(x)$ as x goes to ∞, B is the inflection point of the curve, and K is the rate of change $\left(\frac{\partial f(x)}{\partial x} = -LK\right)$. We use a linear model for LDL, diastolic blood pressure, and HDL so $m_{ldl}()$, $m_{bpd}()$, and $m_{hdl}()$ are linear functions of x.

4.3.2.1 Seemingly unrelated regressions using Normal errors

Because all the MetS risk factors are likely to depend on each other within an individual, we cannot assume that the association between minutes in MVPA and each risk factor is independent. Thus we allow correlation among the errors within an individual. This is often referred to as *Seemingly unrelated regressions* (SUR) (Zellner (1962)). The model for SUR is written as:

$$\begin{aligned} \mathbf{Y_i} &= \mathbf{m}(\mathbf{t_i}, \boldsymbol{\gamma}) + \boldsymbol{\epsilon_i} \\ \boldsymbol{\epsilon_i} &\overset{ind}{\sim} N(\mathbf{0}, \Sigma_m), \end{aligned} \tag{4.25}$$

where $\mathbf{Y_i} = (Y_{i,wst}, Y_{i,glu}, Y_{i,tri}, Y_{i,bps}, Y_{i,bpd}, Y_{i,ldl}, Y_{i,hdl})$, $\mathbf{m}(\cdot)$=($m_{wst}()$, $m_{glu}()$, $m_{tri}()$, $m_{bps}()$, $m_{bpd}()$, $m_{ldl}()$, $m_{hdl}()$), and Σ_m is an unstructured 7×7 covariance matrix. The model in 4.25 is more general than Zellner (1962) since it involves nonlinear functions. Gallant (1975) proposed a method to estimate parameters in the nonlinear case that are consistent and asymptotically normal. William Griffiths wrote a chapter in *Computer-Aided Econometrics* (Giles (2003)) giving an overview of Bayesian inference in SUR for the linear and nonlinear case. More recently, Direct Monte Carlo techniques were proposed by Zellner and Ando (2010) and Ando and Zellner (2010) for efficient Bayesian estimation, but these are restricted to the linear case. The model in 4.25 is somewhat restrictive since it assumes normality of errors.

Based on Figure 4.3, some of the MetS risk factors are highly skewed and thus a symmetric distribution may not be appropriate. Figure 4.6 shows Normal quantile plots of standardized residuals after fitting the model in 4.25. It is apparent that assuming multivariate normality for the errors is not plausible for our application. The estimation method and calculation of standardized residuals are explained in subsequent sections. We considered transformation of response variables, but problems still persisted as there were no good transformations to approximate normality for some variables. Additionally, interpretation is simpler in the original scale, so we prefer a method that does not require transformation all of our response variables.

4.3.2.2 Seemingly unrelated regressions using mixture of Normal errors

Galimberti et al. (2016) proposed using a mixture of normal distributions to model the errors in SUR in the linear case and to use maximum likelihood to estimate error variances and mixing proportions. We extend this model to the nonlinear case and adopt a Bayesian framework. The proposed model uses a mixture of H normal distributions for the seven dimensional error distri-

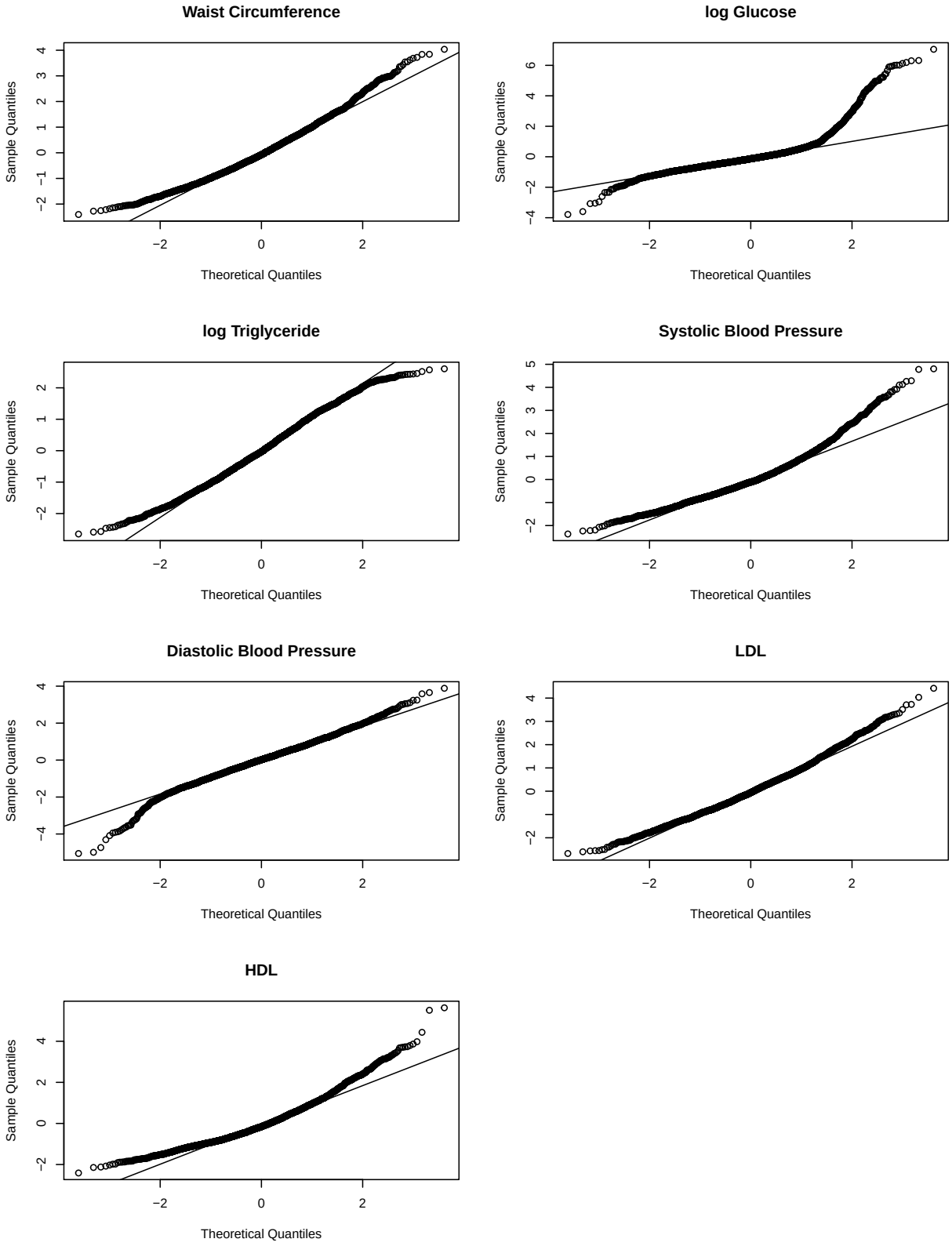

Figure 4.6: Normal quantile plots of standardized residuals from model in 4.25.

bution. The n seven-dimensional error vectors $\boldsymbol{\epsilon_i}$ are assumed to be independent and identically distributed. This model can be written as:

$$\begin{aligned}
&\mathbf{Y_i} = \boldsymbol{\gamma_0} + \mathbf{m}(\mathbf{t_i}, \boldsymbol{\gamma}) + \boldsymbol{\epsilon_i}, \\
&\boldsymbol{\epsilon_i} \overset{ind}{\sim} \sum_{h=1}^{H} p_h N(\mathbf{v_h}, \Sigma_{m,h}), \\
&\text{such that} \\
&\sum_{h=1}^{H} p_h \mathbf{v_h} = \mathbf{0}, \\
&\sum_{h=1}^{H} p_h = 1, \\
&\lambda_h = \boldsymbol{\gamma_0} + \mathbf{v_h}.
\end{aligned} \tag{4.26}$$

In a slight change of notation, $\boldsymbol{\gamma_0}$ represents the overall intercept vector for each function $m()$. The seven-dimensional vector $\mathbf{v_h}$ represents the deviation from the overall mean for the intercept for the hth error component. Although the values $\mathbf{v_h}$ are constrained in 4.26, the component specific intercept terms λ_h are not since none of our intercept terms are constrained. Each value of p_h denotes the weight for each component of the mixture. The 7×7 matrix $\Sigma_{m,h}$ is the covariance matrix for the hth component of the mixture. To write the model in 4.26 in a way that is better suited for MCMC, we introduce an n-dimensional latent vector $\boldsymbol{\zeta}$ whose ith element indicates to which component of the mixture $\mathbf{Y_i}$ belongs. The model is then written as:

$$\begin{aligned}
\mathbf{Y_i}|(\zeta_i = h) &= \lambda_h + \mathbf{m}(\mathbf{t_i}, \boldsymbol{\gamma}) + \boldsymbol{\epsilon_i}, \\
\boldsymbol{\epsilon_i} &\overset{ind}{\sim} N(\mathbf{0}, \Sigma_{m,h}), \\
\zeta_i &\overset{ind}{\sim} \text{Cat}(H, \mathbf{p}),
\end{aligned} \tag{4.27}$$

where $\mathbf{p} = (p_1, p_2, ..., p_H)$ and $\text{Cat}(H, \mathbf{p})$ refers to the categorical distribution with probabilities $\mathbf{p}$ corresponding to H categories. The overall intercept vector can be computed as:

$$\boldsymbol{\gamma_0} = \lambda_1 - \sum_{h=1}^{H} \pi_k(\lambda_1 - \lambda_h). \tag{4.28}$$

4.4 Estimation

We proceed with Bayesian estimation using MCMC for both components of our model. To simplify calculations, we split the estimation into the two components. We first estimate the posterior distributions of the parameters of the measurement error model in 4.15-4.18. Using the output from the MCMC, we calculate values of usual minutes in MVPA, t_i using 4.23 for each individual and each MCMC iteration. Then, using those t_i draws, we fit the regression model in 4.26. We implement the estimation approach on the complete case data as well as on the five imputed data sets. Before providing additional information about the estimation method, we formulate the complete likelihood function for our data and explain our choices of priors.

4.4.1 Complete likelihood

Let $f()$ denote a probability density/mass function. The likelihood for individual i can be written as:

$$L(\boldsymbol{\theta}|\mathbf{W_i}, \mathbf{Y_i}, Z_i) = \int\int f(\mathbf{Y_i}|Z_i, t_i, \boldsymbol{\theta}) f(\mathbf{W_i}|Z_i, b_{1i}, b_{2i}, \boldsymbol{\theta}) f(b_{1i}, b_{2i}|\boldsymbol{\theta}) db_{1i} db_{2i} \tag{4.29}$$

$$= \int\int f(\mathbf{Y_i}|Z_i, b_{1i}, b_{2i}, \boldsymbol{\theta}) \left\{ \prod_{j=1}^{J} f(I(W_{ij} > 0)|Z_i, b_{1i}, \boldsymbol{\theta}) \right\} f(W_{i+}|Z_i, b_{2i}, \boldsymbol{\theta}) \tag{4.30}$$

$$\times f(b_{1i}, b_{2i}|\boldsymbol{\theta}) db_{1i} db_{2i}, \tag{4.31}$$

where $f(\mathbf{Y_i}|Z_i, b_{1i}, b_{2i}, \boldsymbol{\theta})$ is given by 4.26, $f(I(W_{ij} > 0)|Z_i, b_{1i}, \boldsymbol{\theta})$ is given by 4.15, $f(W_{i+}|Z_i, b_{2i}, \boldsymbol{\theta})$ is given by 4.17 and $f(b_{1i}, b_{2i}|\boldsymbol{\theta})$ is given by 4.18. We let $\boldsymbol{\theta} = (\hat{\boldsymbol{\delta}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, \Sigma_b, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{p}, \boldsymbol{\Sigma_m})$. Since we assume independence across individuals, the full likelihood is:

$$L(\boldsymbol{\theta}|\mathbf{W},\mathbf{Y},\mathbf{Z}) = \prod_{i=1}^{n} L(\boldsymbol{\theta}|\mathbf{W_i},\mathbf{Y_i},Z_i). \tag{4.32}$$

4.4.2 Priors

We let all parameters be independent *apriori*, and choose proper priors for all parameters in the model. Most of the priors are non-informative and where possible, conjugate. The priors for the measurement error model parameters are given below:

$$\boldsymbol{\beta} \sim N(\mathbf{0_{12}}, 1000 \times I_{12}) \tag{4.33}$$

$$\boldsymbol{\alpha} \sim N(\mathbf{0_{12}}, 1000 \times I_{12}) \tag{4.34}$$

$$\phi_g \sim \text{Unif}(-1,1), g = 1,2 \tag{4.35}$$

$$\Sigma_b = \boldsymbol{\sigma_b}'\Omega\boldsymbol{\sigma_b} \tag{4.36}$$

$$\Omega \sim \text{LKJ}(1.0) \tag{4.37}$$

$$\sigma_b \sim \text{Cauchy}^{+}(0,1) \tag{4.38}$$

where LKJ(1.0) is a distribution for correlation matrices proposed by Lewandowski et al. (2009). The SUR model has seven components, each with parameter vector $\boldsymbol{\gamma} = (\boldsymbol{\gamma_{wst}}, \boldsymbol{\gamma_{glu}}, \boldsymbol{\gamma_{tri}}, \boldsymbol{\gamma_{bps}}, \gamma_{bpd,1}, \gamma_{ldl,1}, \gamma_{hdl,1})$. The parameter vectors for the nonlinear regressions $\boldsymbol{\gamma_{wst}}, \boldsymbol{\gamma_{glu}}, \boldsymbol{\gamma_{tri}}, \boldsymbol{\gamma_{bps}}$ each have elements L, K, B. For the linear regression components, parameters $\gamma_{bpd,1}, \gamma_{ldl,1}, \gamma_{hdl,1}$ are the slopes associated with minutes in MVPA and each respective MetS risk factors. The intercept vector for the regression is $\boldsymbol{\lambda} = (\boldsymbol{\lambda_{wst}}, \boldsymbol{\lambda_{glu}}, \boldsymbol{\lambda_{tri}}, \boldsymbol{\lambda_{bps}}, \boldsymbol{\lambda_{bpd}}, \boldsymbol{\lambda_{ldl}}, \boldsymbol{\lambda_{hdl}})$, where each $\boldsymbol{\lambda_u}$ is a H dimensional vector. The priors for the regression model parameters are given below:

$$\boldsymbol{\lambda_{wst}} \sim N(\mathbf{98_H}, 17^2 I_H) \tag{4.39}$$

$$\gamma_{wst,L} \sim N(7, 8^2) \tag{4.40}$$

$$\gamma_{wst,K} \sim N(3, 1.5^2) \tag{4.41}$$

$$\gamma_{wst,B} \sim N(2.11, .4^2) \tag{4.42}$$

$$\boldsymbol{\lambda_{bps}} \sim N(\mathbf{130_H}, 7^2 I_H) \tag{4.43}$$

$$\gamma_{bps,L} \sim N(18, 5^2) \tag{4.44}$$

$$\gamma_{bps,K} \sim N(3, 1^2) \tag{4.45}$$

$$\gamma_{bps,B} \sim N(1.3, 1^2) \tag{4.46}$$

$$\boldsymbol{\lambda_{glu}} \sim N(\mathbf{4.7_H}, 1^2 I_H) \tag{4.47}$$

$$\boldsymbol{\lambda_{ldl}} \sim N(\mathbf{0_H}, 100^2 I_H) \tag{4.53}$$

$$\gamma_{ldl,1} \sim N(0, 100^2) \tag{4.54}$$

$$\boldsymbol{\lambda_{tri}} \sim N(\mathbf{4.73_H}, .6^2 I_H) \tag{4.55}$$

$$\gamma_{tri,L} \sim N(.12, .4^2) \tag{4.56}$$

$$\gamma_{tri,K} \sim N(4.88, 2^2) \tag{4.57}$$

$$\gamma_{tri,B} \sim N(2.11, .4^2) \tag{4.58}$$

$$\boldsymbol{\lambda_{hdl}} \sim N(\mathbf{0}, 100^2 I_H) \tag{4.59}$$

$$\gamma_{hdl,1} \sim N(0, 100^2) \tag{4.60}$$

$$\Sigma_{m,h} \sim \text{Inv-Wish}(8, I_{7\times 7}), h = 1, ...H \tag{4.61}$$

$$\mathbf{p} \sim \text{Dirichlet}(\mathbf{1_H}) \tag{4.62}$$

Priors for waist circumference, glucose, triglycerides, and systolic blood pressure regression parameters were chosen to be informative. We used data from the 2001 NHANES to construct these priors. We used the self-report for physical activity where participants categorized their physical activity levels between 1 and 4, 4 being most active. To set prior means for $\boldsymbol{\lambda}$, we took the mean values of the respective MetS risk factors for those individuals who reported the lowest usual physical activity on the self report. Prior means for L were chosen by taking the difference in mean MetS levels for individuals who responded a 4 for physical activity level and those who responded 1. Prior means for B where chosen to be equal to $20^{.25} = 2.11$ since the Physical Activity Guidelines recommend approximately 20 minutes of activity at a 3.0 MET level per day, on average. Prior means for K were selected to be equal to 3, so that there is little additional benefit beyond 80 minutes per day of MVPA with an inflection point at 2. Prior variances were selected using the 2001 NHANES data as well and allowing a small amount of mass at 0 for the priors for K and B. Priors for the regression parameters for LDL, HDL, and diastolic blood pressure are selected to be flat. We selected the prior for the covariance matrix of errors and for the component weights to be conjugate.

4.4.3 Estimation of measurement error model

It was computationally expensive to run the MCMC on the full model at once, so we carried out estimation for the two components of the model in sequence. We used Stan for estimation of parameters in the measurement error model. Stan uses Hamiltonian Monte Carlo to generate draws from posterior distributions. In addition to outputting posterior samples of the parameters for the measurement error model, it also outputs a vector of t values for each individual, one value for each individual for each iteration of the MCMC. We use the t draws for the corresponding individuals who also have data on MetS risk factors in the estimation the regression parameters described in the next section.

4.4.4 Estimation of regression model

To estimate the parameters of the regression model, we wrote our own MCMC in `R` and C++. Although we tried to implement this portion in Stan as well, it was computationally inefficient. We used a Gibbs sampler to generate samples from the joint posterior distribution of all SUR parameters. Full conditional distributions are given in the Appendix. The intercepts for all components of the mixture distribution for the residuals, all covariance matrices of the mixture distribution, and the mixing weights, all had conjugate full conditional distributions. We sampled the remaining regression parameters using a Metropolis step with Normal random walk proposals. Due to the high dimension of the random walk, we allow for a long burn-in period and run the chains for a large number of iterations and thin after. We tune the proposal covariance matrix using the recommended $2.4^2 Var(\theta|y)/d$ in Gelman et al. (2014), pg. 290, where d is the dimension of θ. Because label switching can be an issue with mixture models, we used the method of Stephens (2000) to alleviate this issue. Stephens (2000) uses a decision theoretic approach to address the problem of label switching, where you find the labeling that minimizes the Kullback-Leibler divergence between average component weights for the entire MCMC run and the classfication for each MCMC iteration. This method is implemented in the R package `label.switching` (Papastamoulis (2016)).

4.5 Results

We used the 3336 individuals 18 years or older, who had 6 valid days of at least 10 hours of wear time for estimation of the model in the complete data case; for the imputed data case, we used 7872 individuals. There were 1399 individuals with valid accelerometry data and full MetS risk factor data. After imputing the missing accelerometry data, there were 3305 individuals with complete MetS risk factor data. To approximate the posterior distribution of the measurement error model in Section 4.3.1, we generated eight chains each of length 4000, and used the first 2000 as burn-in in Stan. Trace plots and Gelman-Rubin diagnostics showed good mixing. For the regression model, we generated three chains of length 1,000,000, using the first 250,000 as burn-in, and thinning every 15 iterations to save on memory and reduce auto-correlation. We iterated the chains for the regression model much longer because of the high dimension of the random walk. Starting values for the regression coefficients and error variances of the nonlinear function were obtained by calculating nonlinear least squares estimates using the fourth root of the mean, observed minutes in MVPA for each individual as the explanatory variable. Ordinary least squares estimates were used for the linear function. Starting correlations were set to 0. For the two other chains, 99.99% upper and lower confidence limits were used as starting values of regression coefficients, respectively. Starting variances were taken to be half times and two times the estimates previously mentioned, respectively. The three chains showed good mixing in traceplots and Gelman-Rubin diagnostics close to one (< 1.1) for all parameters. This entire process was done for the complete case data as well as the five imputed data sets.

4.5.1 Number of mixture components

In the estimation of 4.26, the value of H is considered to be fixed. Although it is possible to estimate this quantity in a Bayesian fashion using Reversible Jump MCMC (Green (1995)), we chose a more pragmatic approach given the complexity and high computational costs we already had to this point. To determine H, we used Deviance Information Criterion (DIC) (Spiegelhalter et al. (2002)). In the notation of Spiegelhalter et al. (2002) DIC is defined as:

Table 4.6: DIC for different values of number of mixture components H for complete case and imputed data.

H	DIC Complete	DIC Imputed
1	60159	144344
2	58394	139587
3	57771	137989
4	57816	136792
5		135787
6		135944

$$\begin{aligned} DIC &= \overline{D(\theta)} + p_D \\ &= \overline{D(\theta)} + (\overline{D(\theta)} - D(\tilde{\theta})) \\ &= 2\overline{D(\theta)} - D(\tilde{\theta}) \\ &= -4E_\theta\left(\log f(\mathbf{y}|\theta)|\mathbf{y}\right) + 2\log f(\mathbf{y}|\tilde{\theta}), \end{aligned} \tag{4.63}$$

where we let $\tilde{\theta}$ be a posterior mean. Table 4.6 shows DIC values for both the complete case data and imputed data under multiple values of H. Since smaller DIC reflects a better fitting model, the optimal number of mixing components for the complete case analysis is three, and for the imputed data case is five. When $H = 1$, the model in 4.26 reduces to the model in 4.25.

4.5.2 Estimating the mean function in the SUR

Regression parameter estimates for the nonlinear and linear functions under the model that is fit to the complete data are given in Tables 4.7 and 4.8, respectively, together with a 95% credible set. Letting $\hat{t}_i = \frac{1}{M}\sum_{m=1}^{M} t_i^{(m)}$ be the estimated usual minutes of MVPA for individual i based on M MCMC simulations, we plot $\hat{t}_i$ versus observed MetS risk factor and overlay the fitted mean function in Figure 4.7. Recall that the parameters of interest in the nonlinear function are L and B, where L is the benefit that can be gained from participating in MVPA and B is the inflection point, which can be thought of as the minimum number of minutes of MVPA required to maximize benefits toward the specific MetS risk factor; there are increasing returns for MVPA up

until that point. With increased MVPA, waist circumference can decrease by over 11cm, systolic blood pressure drops by 13 mm Hg, glucose levels drop by more than 13 mg/dL ($e^{\gamma_0} - e^{\gamma_0 - L}$), and triglyceride levels drop by 38 mg/dL, on average. All of these effects are statistically significant and practically important. These results are for when minutes of MVPA approaches about 80 minutes a day, an ambitious goal for most people. However, the inflection points B range in value on the transformed scale from 1.5 for Glucose to 2.3 for triglycerides for fourth root of minutes, which translates to about 5 minutes to 28 minutes of MVPA per day to observe a meaningful change in risk factors. This is a large range to achieve the optimal tradeoff between health benefits and time. These results suggest that a one size fits all approach to exercise guidelines may not be justifiable.

The estimated slopes in the linear regressions are largely in line with our expectations. The coefficients for LDL and HDL are negative but not significantly so. We would expect LDL to decrease with increased MVPA and HDL to increase with increased MVPA, in general. The only unexpected result was a positive slope for diastolic blood pressure suggesting that more physical activity is associated with higher blood pressure instead. While the slope was statistically significant, the estimated increase of 1.44 is modest.

Often of interest is the probability that an individual will have a high level of a particular MetS risk factor for a given usual amount of time in MVPA. In order to calculate this probability we simulate data from the posterior predictive distribution of $\mathbf{Y}$ for the draws of t that we obtained in the estimation step. We then we calculate the proportion of predicted observations that are higher than the critical thresholds for each risk factor. We calculate this probability by simulating 200 replicate data sets of $\mathbf{Y}$ for values of t between 0 and 60 by one minute increments. Figure 4.8 shows these probabilities for each MetS risk factor as a function of minutes in MVPA. We see a fast drop in probability for both systolic blood pressure and glucose, as suggested by the values of parameter estimates. It is also apparent that the probabilities of low HDL and high diastolic blood pressure are actually increasing, which is again surprising. Notice that even with 60 minutes of MVPA a day on average, there is still a 43% chance of having a large waistline. This suggests

that factors other than exercise are likely to have a major impact on one's weight and waistline as well as exercise.

Table 4.7: Parameter estimates in the nonlinear functions for complete case data.

Parameter	MetS RF	Post Mean Estimate	95% Credible Interval
γ_0	Waist (cm)	101.361	(99.283,104.009)
	Glucose (log mg/dL)	4.687	(4.647,4.748)
	Triglyceride (log mg/dL)	4.852	(4.78,4.965)
	Sys Blood Press (mm Hg)	131.47	(128.611,134.533)
L	Waist	11.422	(6.738,19.581)
	Glucose	0.134	(0.083,0.216)
	Triglyceride	0.352	(0.173,0.666)
	Sys Blood Press	13.35	(9.713,17.259)
K	Waist	2.86	(1.379,4.886)
	Glucose	2.46	(1.108,4.273)
	Triglyceride	2.605	(1.095,4.786)
	Sys Blood Press	3.899	(2.35,6.163)
B	Waist	2.211	(1.799,2.786)
	Glucose	1.556	(1.119,1.995)
	Triglyceride	2.332	(1.764,3.047)
	Sys Blood Press	1.681	(1.48,1.876)

Table 4.8: Parameter estimates in the linear functions for complete case data.

Parameter	MetS RF	Post Mean Estimate	95% Credible Interval
γ_0	LDL (mg/dL)	118.134	(110.828,125.603)
	Dias Blood Press (mm Hg)	66.956	(64.608,69.22)
	HDL (mg/dL)	58.14	(55.26,60.958)
γ_1	LDL	-0.385	(-4.075,3.2)
	Dias Blood Press	1.447	(0.346,2.586)
	HDL	-1.171	(-2.53,0.229)

4.5.3 Probability of MetS

MetS is diagnosed when a person exhibits multiple elevated risk factors. Any combination of three elevated factors typically results in a MetS diagnosis. We also calculated the probabilities that an individual exhibits R or more risk factors elevated levels, for R=1,2,...,6. Figure 4.9 shows

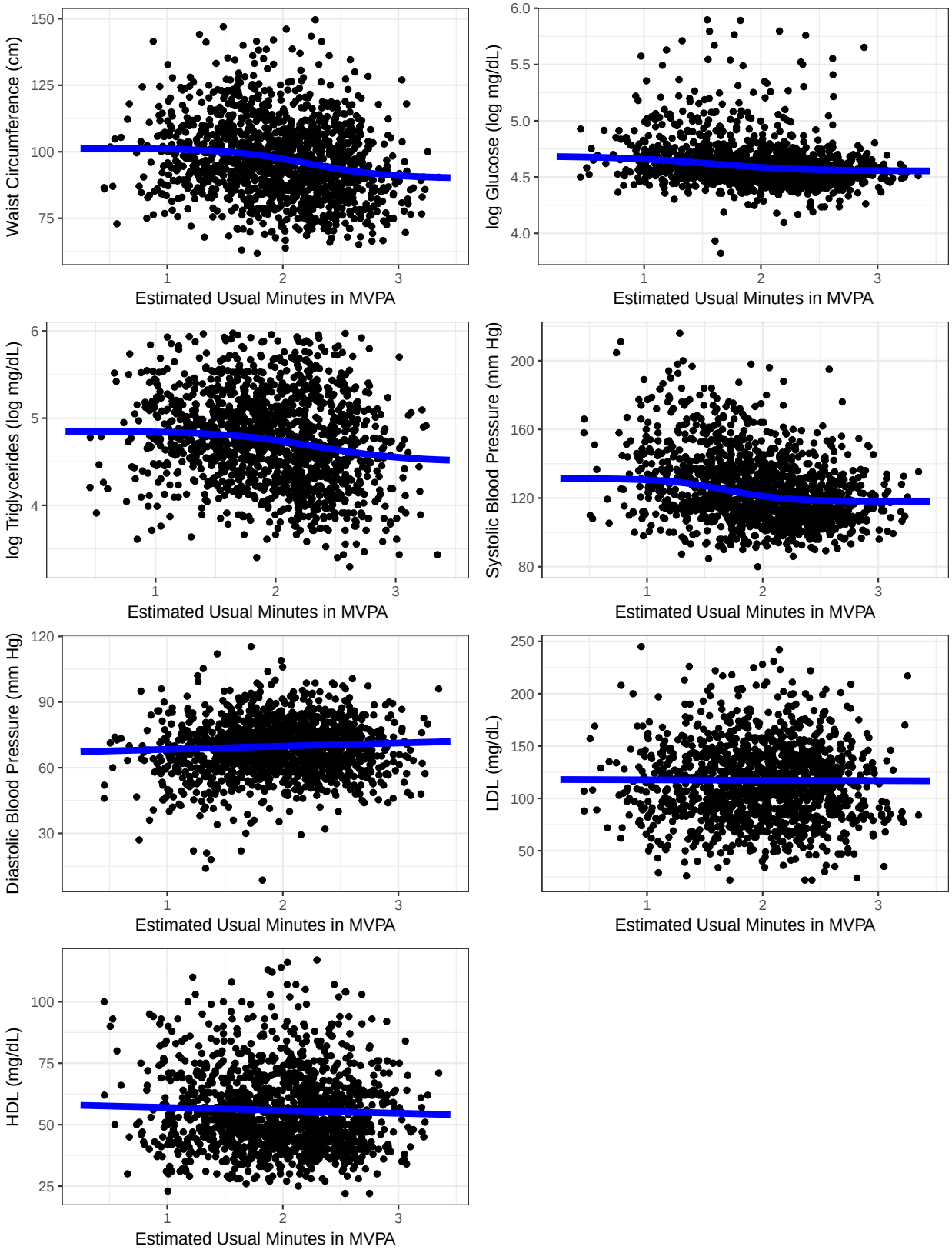

Figure 4.7: Fitted regression functions against $\hat{t}_i$ values.

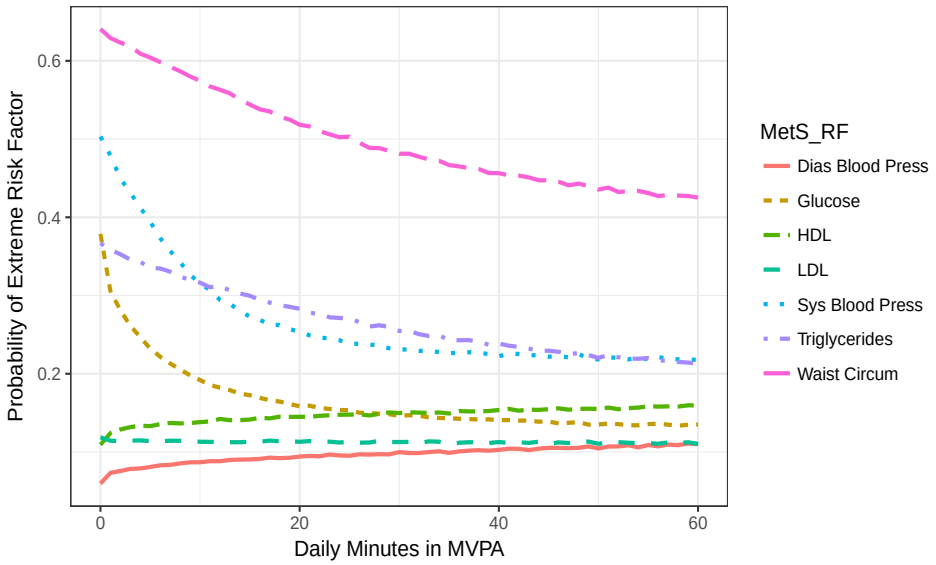

Figure 4.8: Probability of high levels for each MetS risk factor as function of daily minutes in MVPA.

the results, along with pointwise 95% credible intervals. The standard definition of MetS is elevated levels of three or more risk factors, so Figure 4.9 indicates that the average probability of MetS diagnosis for those participating in zero minutes of MVPA is 40% while for those participating in 20 minutes is 23%. It is also interesting to note that even for those who participate in 60 minutes of MVPA daily on average, there is still a nearly 70% chance of having at least one elevated MetS risk factor. This suggests that there is more to MetS risk factors than time in MVPA.

4.5.4 Residual analysis

To assess the fit of model 4.26, we calculate standardized residuals and perform residual analysis. We let the residual vector for individual i be:

$$\mathbf{r_i} = \frac{\mathbf{Y_i} - E(\tilde{\mathbf{Y}}_\mathbf{i}|\mathbf{Y_i})}{\sqrt{var(\tilde{\mathbf{Y}}_\mathbf{i}|\mathbf{Y_i})}}, \tag{4.64}$$

where $f(\tilde{\mathbf{Y}}_\mathbf{i}|\mathbf{Y_i})$ denotes the posterior predictive distribution which is defined as:

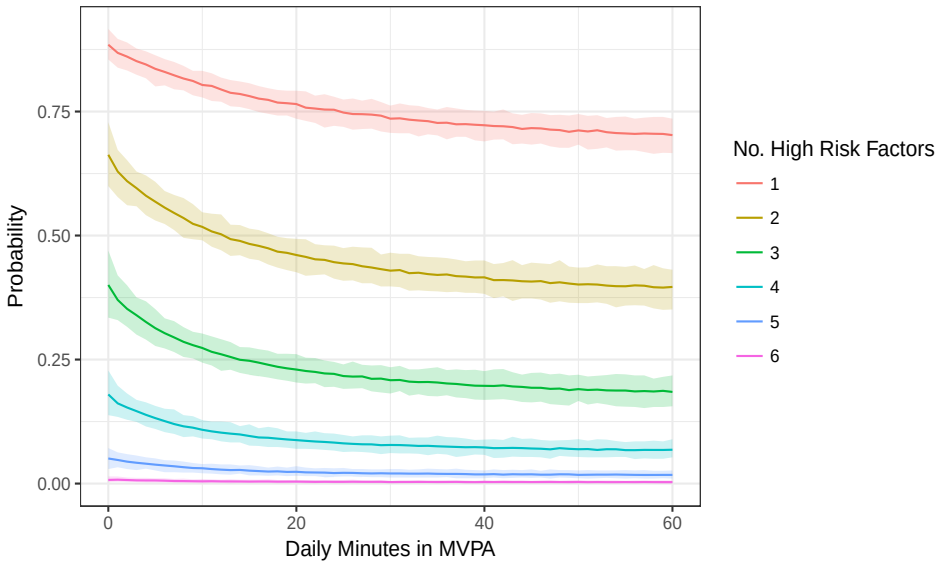

Figure 4.9: Probability of exhibiting $R = 1, 2, ...6$ elevated levels of MetS risk factor as a function of daily minutes in MVPA.

$$f(\tilde{\mathbf{Y}}|\mathbf{Y}, \mathbf{W}) = \int\int\int f(\tilde{\mathbf{Y}}|\mathbf{Z}, \mathbf{b_1}, \mathbf{b_2}, \boldsymbol{\theta}) f(\mathbf{b_1}, \mathbf{b_2}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{W}) d\mathbf{b_1} d\mathbf{b_2} d\boldsymbol{\theta}, \tag{4.65}$$

for $f(\mathbf{b_1}, \mathbf{b_2}|\boldsymbol{\theta})$ and $f(\tilde{\mathbf{Y}}|\mathbf{Z}, \mathbf{b_1}, \mathbf{b_2}, \boldsymbol{\theta})$ as defined in 4.18 and 4.26. We calculate the expectation and variance in 4.64 using a Monte Carlo simulation. We simulated 1000 replicate data sets from 4.65 by taking 1000 draws from the posterior distribution of $\boldsymbol{\theta}$, using those values to simulate 1000 pairs of $(\mathbf{b_1}, \mathbf{b_2})$ for each individual, and then using $\boldsymbol{\theta}$, $(\mathbf{b_1}, \mathbf{b_2})$, and the model matrix $\mathbf{Z}$ to simulate new observations from the distribution $f(\tilde{\mathbf{Y}}|\mathbf{Z}, \mathbf{b_1}, \mathbf{b_2}, \boldsymbol{\theta})$, in 4.26.

We plot $\hat{t}_i$ versus $\hat{\mathbf{r}}_\mathbf{i}$ for each MetS risk factor in Figure 4.10 for the complete case analysis. None of the residual plots for the MetS risk factors appear to show nonconstant variance or other trends, when plotted against the predicted ususal minutes of MVPA.

Using posterior means of mixture weights p and mixture specific intercept deviations $\mathbf{v}$, we can calculate the density functions of the error distributions for 4.26. Marginal error densities for each MetS risk factor are plotted in Figure 4.11. In these densities, we can see signs of non-normality.

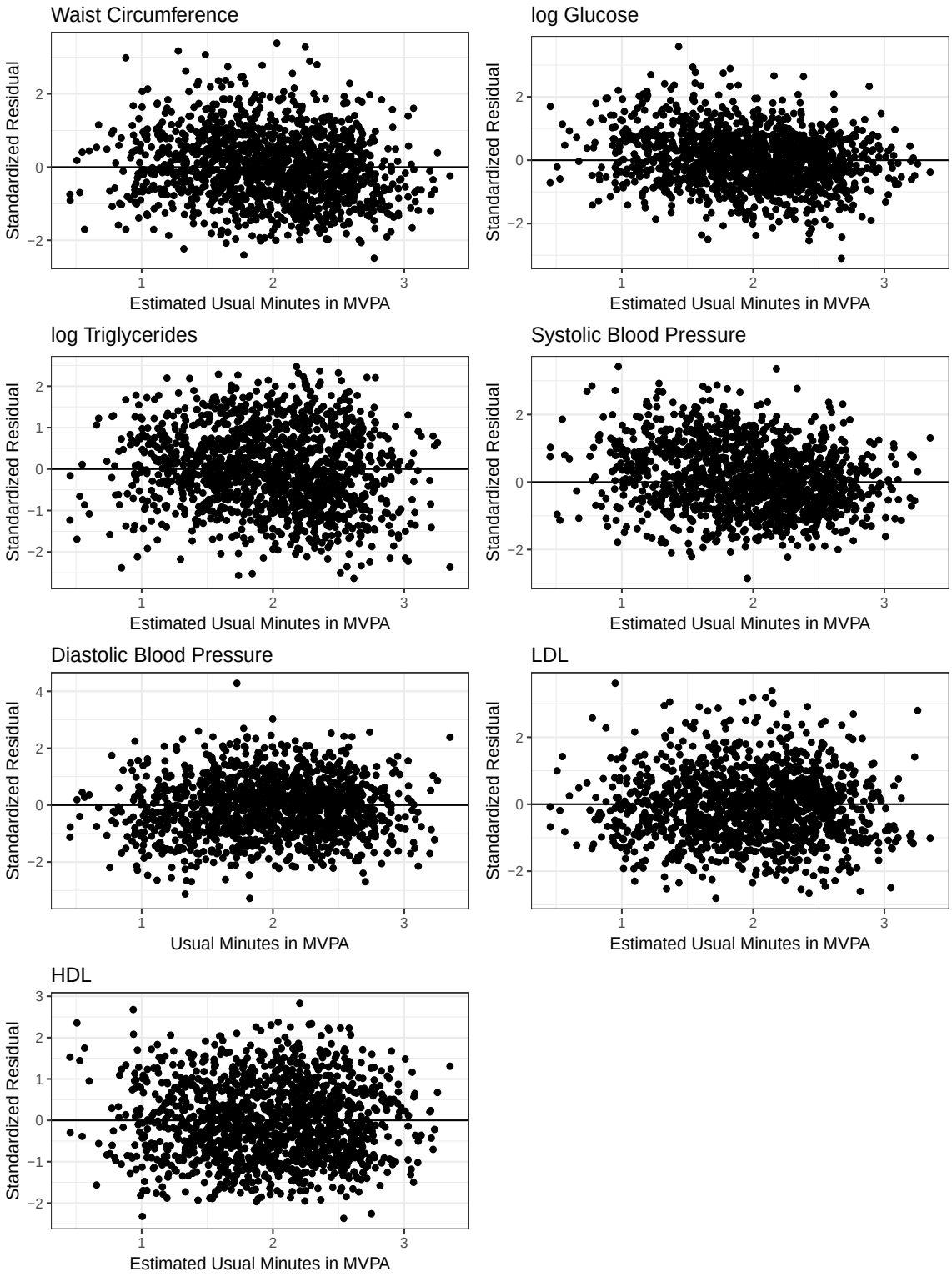

Figure 4.10: Plots of standardized residuals against $\hat{t}_i$ for the model fitted to the complete case data. Residuals are computed as in 4.64.

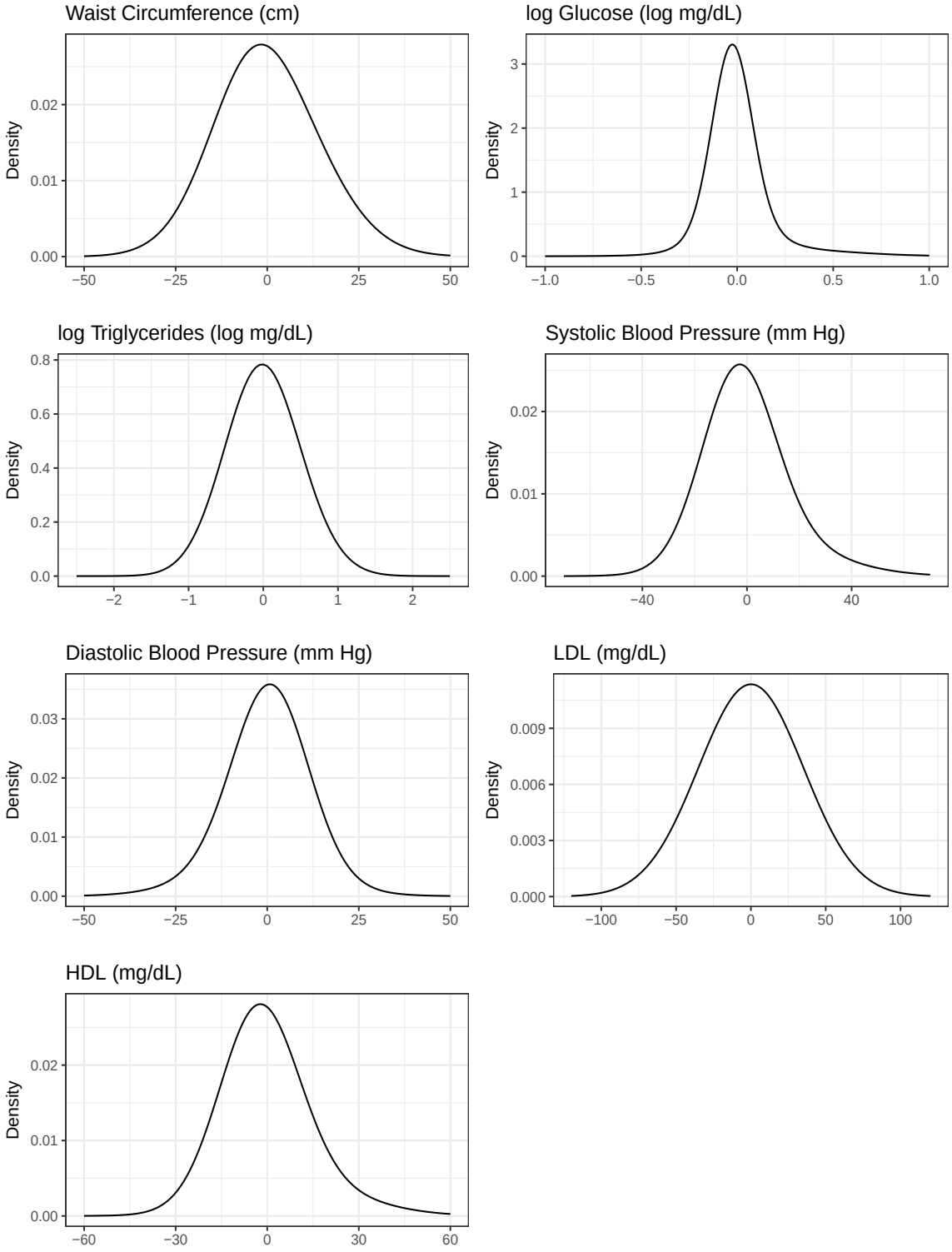

Figure 4.11: Marginal error densities from model in 4.26.

4.5.5 Component analysis

For the complete case analysis $H = 3$. We classified each individual into one of those three groups based on each individuals posterior mode for ζ_i. Creating 2-way contingency tables based on groupings and demographic variables, we calculated chi-square tests for proportions, and the results suggested that the components tend to correspond to gender, age, and BMI (all p<0.0001), while education and race did not appear to affect the clustering ($p = 0.81, p = 0.11$, respectively). The first component included more males, in the 35-65 age range, and with high BMI. The second component corresponded to slightly younger persons with BMI in the healthy to slightly overweight range. The third component was more female, older (50+), and with the lowest BMI.

Table 4.9 shows the estimates of λ and p for each risk factor for each of the three components fit to the complete case data. We see that the first component included 10% of the individuals in the study, the second component included 71% and the remaining 19% were allocated to the third component. In addition individuals in the first cluster tend to be more unhealthy based on MetS risk factors alone, as they have the highest waist circumference, glucose, triglycerides, and diastolic blood pressure levels along with the lowest HDL levels in addition to a high systolic blood pressure level. This is consistent with the fact that the first component has the highest BMI levels. The third component has more females and more elderly persons, who tend to have smaller waist circumference. Additionally, women tend to have lower LDL and higher HDL levels, which aligns with the grouping.

Table 4.10 shows estimates of standard deviations and correlations for $\Sigma_{m,h}$ for each of the three components for the complete case analysis. For the second component, where a majority of individuals are classified, we see that all the correlations are positive except those with HDL. This is what we would expect, because in general, elevated levels of all the MetS risk factors are not desirable, except for HDL, where high levels are generally good. There are some negative correlations not related to HDL in components one and three, but most are negligible. All of the non-negligible negative correlations not including HDL are in component one, between waist and

LDL, triglycerides and LDL, and triglycerides and diastolic blood pressure. Correlations were closer to zero for the most part in component one compared to the other components.

Table 4.9: Intercept estimates under each 3 components fit to complete case data.

h	p	λ_{wst}	λ_{glu}	λ_{tri}	λ_{bps}	λ_{ldl}	λ_{bpd}	λ_{hdl}
1	0.10	112.46	4.90	5.24	133.50	114.17	71.13	49.83
2	0.71	101.28	4.66	4.79	127.72	120.59	67.24	56.31
3	0.19	95.75	4.68	4.88	144.48	111.05	63.65	69.51

4.5.6 Imputed data results

We fit the same model to each of the five data sets with imputed missing values following the same approach in the complete case data. The parameters of interest are the regression coefficients, which are shown in Tables 4.11 and 4.12 for the nonlinear and linear regressions, respectively. Results in Table 4.11 are similar to those given in Table 4.7. Parameters γ_0, L, K have overlapping 95% credible intervals when based on the complete or the imputed datasets. There are some differences in the B parameter, however. We find that the minutes in MVPA required to reach the inflection point where returns in terms health are no longer increasing is smaller. The differences are more noticeable in the linear regression paraeters, see Tables 4.12 and 4.8. Although the intercepts are similar, the estimates differ. The slope for LDL is negative and significant for the imputed data while it was not different from zero in the complete case data. This is what we would expect, because it suggests greater benefits for LDL levels given more time in MVPA. The slope estimate for diastolic blood pressure is not significantly different from zero in the imputed data case, whereas it was positive in the complete case data. Again, this is more interpretable since we do not typically associate more MVPA with higher blood pressure levels. The slope estimate for HDL was not significantly different from zero in the complete case data, but it is significantly less than zero in the imputed data case. This is surprising since HDL is considered the "good" cholesterol. These differences are attributed to both the imputation mechanism as well as the additional observations of MetS risk factors that could be added as a result of the imputation. Figure 4.12 shows the

Table 4.10: Standard deviation and correlation estimates for each of the three components of the mixture for the complete case analysis.

	wst	glu	tri	bps	ldl	bpd	hdl
wst	15.07	-0.019	0.056	0.123	-0.231	0.244	-0.26
glu		0.361	-0.083	0.118	0.03	0.001	0.04
tri			0.493	-0.041	-0.252	-0.199	-0.47
bps				16.038	-0.057	0.461	0.10
ldl					36.068	0.26	0.18
bpd						9.183	-0.04
hdl							11.44
	wst	glu	tri	bps	ldl	bpd	hdl
wst	14.615	0.358	0.32	0.255	0.141	0.196	-0.28
glu		0.102	0.185	0.305	0.04	0.136	-0.16
tri			0.486	0.15	0.34	0.112	-0.46
bps				13.593	0.117	0.485	-0.08
ldl					34.775	0.169	-0.01
bpd						10.296	-0.02
hdl							12.79
	wst	glu	tri	bps	ldl	bpd	hdl
wst	10.374	0.209	0.429	0.178	0.054	-0.056	-0.44
glu		0.175	0.192	0.135	-0.057	-0.02	-0.30
tri			0.516	0.009	-0.072	-0.035	-0.48
bps				23.861	0.054	0.349	-0.31
ldl					34.88	0.025	0.05
bpd						17.174	0.06
hdl							21.21

probabilities of having elevated levels of each of the MetS risk factor as a function of usual time in MVPA. The results are similar to those shown in Figure 4.8, except for the steeper drop in probability for glucose, triglycerides, and systolic blood pressure due to the larger estimates of B.

4.6 Discussion

We explored the relationships between minutes in MVPA adjusted for measurement error, and seven MetS risk factors. We used data from 2003-2006 NHANES and performed two analyses, one for the complete case data (N=3337) where individuals provided six days of accelerometer

Table 4.11: Parameter estimates in the nonlinear functions for imputed data.

Parameter	MetS RF	Post Mean Estimate	95% Credible Interval
γ_0	Waist (cm)	104.313	(91.215,112.78)
	Glucose (log mg/dL)	4.763	(4.618,4.979)
	Triglyceride (log mg/dL)	4.855	(4.531,5.264)
	Sys Blood Press (mm Hg)	131.204	(124.759,139.2)
L	Waist	12.749	(9.207,18.545)
	Glucose	0.129	(0.086,0.193)
	Triglyceride	0.264	(0.146,0.457)
	Sys Blood Press	12.987	(10.542,16.352)
K	Waist	2.637	(1.623,3.985)
	Glucose	2.967	(1.745,4.644)
	Triglyceride	2.832	(1.17,4.93)
	Sys Blood Press	4.949	(3.586,6.44)
B	Waist	2.181	(1.939,2.517)
	Glucose	1.289	(0.99,1.552)
	Triglyceride	1.794	(1.416,2.176)
	Sys Blood Press	1.426	(1.301,1.534)

measurements, and one where we performed multiple imputation to fill in missing times in the accelerometer information (N=7873). We used a measurement error approach that assumes that the measurements from the accelerometer are unbiased for an individuals' usual physical activity level, and the accelerometer records MVPA if and only if the individual is actually participating in MVPA. Our model for the accelerometer data takes into consideration the autocorrelation present due to the fact that data are collected six days in a row, and we also model the variance of minutes in MVPA as a function of age using a cubic polynomial regression. Transforming the accelerometry data to its fourth root makes the assumption of normality a reasonable assumption, and decreases dependence between the measurement error variance and time in MVPA. We used a Taylor approximation to calculate measurement error-adjusted *usual* values of MVPA for each individual. We defined a nonlinear mean function relating usual time in MVPA to four of the MetS risk factors and used a linear function for the remaining three. Because all seven of the MetS risk factors are correlated for an individual, we opted for a seemingly unrelated regressions approach to the problem. We showed that normality is not a reasonable assumption for the errors in the

Table 4.12: Parameter estimates in the linear functions for imputed data.

Parameter	MetS RF	Post Mean Estimate	95% Credible Interval
γ_0	LDL (mg/dL)	121.84	(112.982,139.546)
	Dias Blood Press (mm Hg)	69.119	(65.356,75.981)
	HDL (mg/dL)	60.369	(51.076,69.821)
γ_1	LDL	-6.035	(-8.368,-3.606)
	Dias Blood Press	0.098	(-0.699,0.894)
	HDL	-2.205	(-3.247,-1.293)

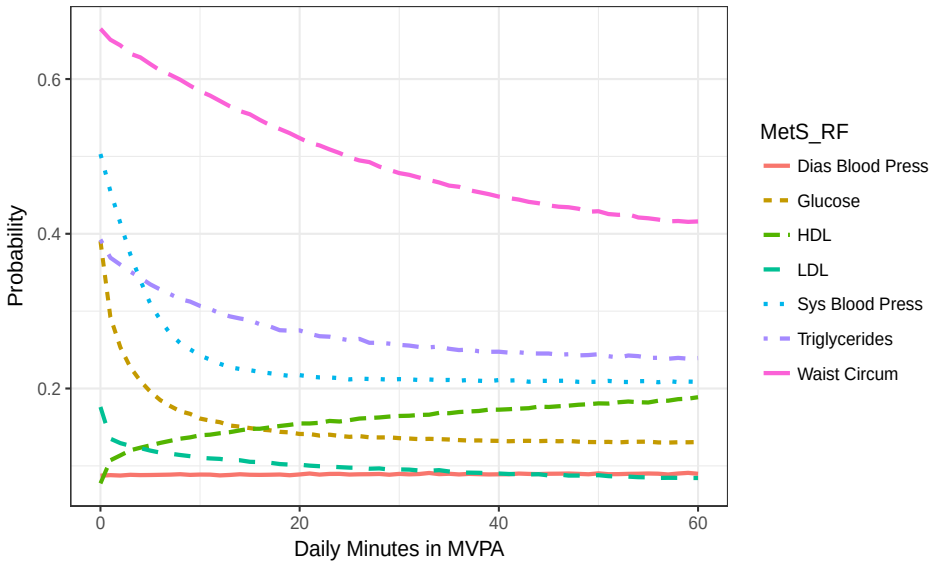

Figure 4.12: Probability of high levels for each MetS risk factor as function of daily minutes in MVPA for the imputed data case.

SUR, so instead we used a mixture of multivariate normal distributions and selected the number of mixing components using DIC.

We carried out estimation in this problem within the Bayesian framework, and used Stan to estimate the parameters in the measurement error model. We wrote our own MCMC for estimation in the regression model. We constructed informative priors for the regression coefficients of the nonlinear mean functions using data from 2001 NHANES, but the remaining priors were chosen to be relatively uninformative, and in the regression model MCMC, conjugate, if possible. Results were consistent for the most part with our expectations and suggest that time in MVPA can help

reduce the risk of developing MetS. Our results indicate that as little as 6 minutes of MVPA a day can help reduce glucose levels by 6.5 mg/dL or 20 minutes of MVPA can reduce one's waistline by 5cm. Additional benefits can be attained with more time spent in MVPA as shown in Figure 4.7. Our results for the linear models were a little more surprising; in the complete case analysis, results suggest that time in MVPA has no effect on LDL, a positive effect on diastolic blood pressure, and a negative effect on HDL. However, using the imputed data, results were slightly different and LDL was found to be significantly negatively associated with time in MVPA and diastolic blood pressure was found to have no relationship with time in MVPA. These, along with quicker responses to MVPA for glucose, triglycerides, and systolic blood pressure were the only major differences between the complete case analysis and the imputed data analysis which suggests that the missing data mechanism is not informative. It is worth noting that doing a linear regression on diastolic blood pressure (HDL) on mean, fourth root observed time in MVPA by individual, results in positive (negative) coefficients as well, even though it goes against current knowledge on these topics[2]. There are at least four possible reasons for this. First, NHANES is a cross sectional dataset and no "treatment" of time in MVPA is being applied. Second, we have not accounted for individulas taking medicine for either blood pressure or cholesterol. NHANES does ask if you are taking a prescription for hypertension, but 73% are missing values, and NHANES only asks if you are told to take a prescription for cholesterol and 78% are missing values. Third, individuals in the sample could be on a physical activity intervention, where they started increasing physical activity levels, but the long run benefits have not been achieved yet. Finally, we only consider total minutes in MVPA, whereas the PAG only counts minutes of MVPA accrued in at least 10 minute bouts. We also note that there have been different findings based on whether one is participating in aerobic or resistance based exercise[3], of which both could be at the moderate to vigorous intensity, and our data does not distinguish between the two. Most of these issues arise due to the data collection mechanism, and we cannot make adjustments once the data are on hand. Nonetheless, residual analysis showed that the model fit well.

[2] https://health.gov/paguidelines/report/pdf/CommitteeReport.pdf
[3] See footone 2

CHAPTER 5. SUMMARY

Regular physical activity is necessary for a healthy and well rounded life. There have been many studies looking at the associations between physical activity and various health outcomes. These studies look at different exercise regimes, intensities, durations, types, etc. across all different demographic groups and populations. These studies are used as the bases for physical activity recommendations for children, adults, and elderly. They are use to help create physical activity interventions for those at risk for serious conditions like heart disease and diabetes, where the physical activity component is often one part of a bigger lifestyle change. As the discipline goes deeper, we need to remember the data used to make these conclusions are measured with error, often including device measurement error, day to day variation, as well as person biases. These measurement errors can have a large impact on the final results and so care must be taken to account for this measurement error either in terms of calibration or directly modeling the error. Even as the technology for objective measures of physical activity moves forward, this will always be an issue to consider as we've shown.

In Chapter 2, we have proposed a general methodology for the measurement of energy balance. We jointly model EE and ΔES measurements in order to borrow information between the two through their relationship via the energy balance principle. Given gold standard measurements for both, we can estimate the biases in less expensive measurements and give calibration equations for these instruments.

In Chapter 3, we developed a Bayesian approach for assessing physical activity of at least moderate intensity occurring in at least 10 minute bouts. The main purpose of estimating the distribution of this activity level was to assess the compliance to the PAG. We developed a two-part model that adjusts physical activity data for measurement error. There were mixed results as PAMS indicated that upwards of 78% of the adult Iowa four county population met the PAG, while only

16% of adults Americans met the guidelines. The discrepancies between these estimates could be attributed to various causes, such as different populations of interest, different accelerometers, and difference in years. Nonetheless, we showed our model fits the data well using posterior predictive model assessment.

In Chapter 4, we modeled the relationship between usual minutes in MVPA per day and the levels of seven MetS risk factors. First, we adjusted observed measurements from accelerometers for measurement error, and then used those adjusted values in a non-linear SUR framework with errors modeled as a mixture of normals. We concluded that an increase in time spent in MVPA is associated with a significant decrease in the probability of having MetS. Additionally, we showed the complex relationships between time in MVPA and levels for each risk factor as some risk factors required small amounts of MVPA for a large benefit while others required larger amounts of MVPA.

5.1 Suggestions for Future Work

We have several suggestions to continue research in each of the three chapters presented. The model in Chapter 2 doesn't account for non-constant variance in the measurement error model, which based on the EBS data is a strong assumption. It would also be interesting to compare our RJMCMC to a fully Bayesian implementation. Additionally, we are not aware of any data sets that have the required measurements for this model. Although we believe it would be a beneficial endeavour to design a study to collect the required data, as previously discussed, it would would carry a high price tag. In both Chapter 3 and 4, we made the assumption that the accelerometer provided data that was unbiased to the truth. As discussed at the end of Chapter 3, this may not be the best assumption, and better measurement error models to deal with accelerometry data are needed. An extension to our work in Chapter 4 would be to do a similar analysis using bouted data and borrowing the measurement error model from Chapter 3. We considered this when we began work on Chapter 4, but with the 2003-2006 NHANES accelerometery data, over 85% of individuals participated in zero 10 minute bouts of activity in at least 3 METs. As far as we know, there are

not data sets as rich as NHANES that have both accelerometry data as well as MetS risk factor data.

BIBLIOGRAPHY

Ainsworth, B., Haskell, W., Herrmann, S., Meckes, N., Bassett, J. D., Tudor-Locke, C., Greer, J., Vezina, J., Whitt-Glover, M., and Leon, A. (2011). Compendium of physical activities: a second update of codes and met values. *Medicine and Science in Sports and Exercise*, 43:1575–81.

Ando, T. and Zellner, A. (2010). Hierarchical bayesian analysis of the seemingly unrelated regression and simultaneous equations models using a combination of direct monte carlo and importance sampling techniques. *Bayesian Analysis*, 5:65–96.

Bai, J., Goldsmith, J., Caffo, B., Glass, T., and Crainiceanu, C. (2012). Movelets: A dictionary of movement. *Electronic Journal of Statistics*, 6:559–78.

Bai, J., He, B., Shou, H., Zipunnikov, V., Glass, T., and Crainiceanu, C. (2014). Normalization and extraction of interpretable metrics from raw accelerometry data. *Biostatistics*, 15:102–16.

Baracos, V. e. a. (2011). Advances in the science and application of body composition measurement. *Journal of Parenteral and Enteral Nutrition*, 36:96–107.

Bassett, D. e. a. (2000). Validity of four motion sensors in measuring moderate intensity physical activity. *Med Sci Sport Exer*, 32:S471–S480.

Berry, S. M., Carroll, R. J., and Ruppert, D. (2002). Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association*, 97:160–69.

Beyler, N., Fuller, W., Nusser, S., and Welk, G. (2015). Predicting objective physical activity from selfreport surveys: a model validation study using estimated generalized least-squares regression. *Journal of Applied Statistics*, 42:555–65.

Bouten, C., Verboeket-van de Venne, W., Westerterp, K., Verduin, M., and Janssen, J. (1996). Daily physical activity assessment: comparison between movement registration and doubly labeled water. *Journal of Applied Physiology*, 81:1019–26.

Bowker, A. H. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association*, 43:572–4.

Butcher, K., Sallis, J. F., Mayer, J. A., and Woodruff, S. (2008). Correlates of physical activity guideline compliance for adolescents in 100 u.s. cities. *Journal of Adolescent Health*, 42:360–368.

Calabro, A. M., Lee, J.-M., Saint-Maurice, P. F., Yoo, H., and Welk, G. J. (2014). Validity of physical activity monitors for assessing lower intensity activity in adults. *Int J Behav Nutr Phys Act*, 11:1–9.

Camhi, S. M., Sisson, S. B., Johnson, W. D., Katzmarzyk, P. T., and Tudor-Locke, C. (2011). Accelerometer-determined moderate intensity lifestyle activity and cardiometabolic health. *Preventive Medicine*, 52:358–60.

Caroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceani, C. M. (2006). *Measurement Error in Nonlinear Models*. Chapman and Hall/CRC.

Casiraghi, F., Lertwattanarak, R., Luzi, L., Chavez, A., Davalli, A., Naegelin, T., Comuzzie, A., Frost, P., Musi, and Folli, F. (2013). Energy expenditure evaluation in humans and non-human primates by sensewear armband. validation of energy expenditure evaluation by sensewear armband by direct comparison with indirect calorimetry. *PLOS One*, 8:1–8.

Catellier, D. J., Hannan, P. J., Murray, D. M., Addy, C. L., Conway, T. L., Yang, S., and Rice, J. C. (2005). Imputation of missing data when measuring physical activity by accelerometry. *Medicine and Science in Sports and Exercise*, pages 555–62.

CDC. National health and nutrition examination survey questionnaire (or examination protocol, or laboratory protocol). hyattsville, md: U.s. department of health and human services, centers for disease control and prevention, 2003-2006.

CDC (2017). Cdc website. https://www.cdc.gov/nchs/fastats/obesity-overweight.htm. Accessed: 2017-01-15.

Consul, P. (1988). *Generalized Poisson Distributions*. CRC Press.

Crouter, S., Churilla, J., and Bassett, D. (2006). Estimating energy expenditure using accelerometers. *European Journal of Applied Physiology*, 98:601–12.

Czado, C., Erhardt, V., Min, A., and Wagner, S. (2007). Zero-inflated generalized poisson models with regression effects on the mean, dispersion and zero-inflation level applied to patent outsourcing rates. *Statistical Modeling*, 7:125–53.

Dalacorte, R., Reichert, C. L., and Vieira, J. L. (2009). Metabolic snydrome and physical activity in southern brazilian community-dwelling elders: a population-based, cross sectional study. *BMC Public Health*, 9:1–8.

Dannecker, K. L. e. a. (2013). A comparison of energy expenditure estimation of several physical activity monitors. *Med Sci Sports Exerc*, 45:2105–12.

de Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer.

Denison, D., Mallick, B., and Smith, A. (1998). Automatic bayesian curve fitting. *Journal of the Royal Statistical Society*, 60:333–50.

DiMatteo, I., Genovese, C. R., and Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika*, 88:1055–71.

Drenowatz, C. and Eisenmann, J. (2011). Validation of the sensewear armband at high intensity exercise. *Eur J Appl Physiol*, 111:883–7.

Fan, R., Chen, V., Xie, Y., Yin, L., Kim, S., Albert, P., and Simons-Morton, B. (2015). A functional data analysis approach for circadian patterns of activity of teenage girls. *Journal of Circadian Rhythms*, 13:1–13.

Fernandez, C. and Steel, M. F. J. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93:359–71.

Ford, E. S., Kohl, H. W., Mokdad, A. H., and Ajani, U. A. (2005). Sedentary behavior, physical activity, and the metabolic syndrome among u.s. adults. *Obesity Research*, 13:608–14.

Franks, P. W., Ekellund, U., Brage, S., Wong, M.-Y., and Wareham, N. J. (2004). Does the association of habitual physical activity with the metabolic syndrome differ by level of cardiorespiratory fitness? *Diabetes Care*, 27:1187–93.

Freedson, P., Melanson, E., and Sirard, J. (1998). Calibration of the computer science and applications, inc. accelerometer. *Med Sci Sports Exercise*, 30:777–81.

Fuller, W. A. (1987). *Measurement Error Models*. New York: John Wiley and Sons.

Galimberti, G., Scardovi, E., and Soffritti, G. (2016). Using mixtures in seemingly unrelated linear regression models with non-normal errors. *Stat Comput*, 26:1025–38.

Gallant, R. A. (1975). Seemingly unrelated nonlinear regressions. *Journal of Econometrics*, 3:35–50.

Gelman, A. (2007). Struggles with survey weighting and regression modeling. *Statistical Science*, 22:153–64.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*. Chapman and Hall/CRC.

Giles, D. E. (2003). *Copmuter-Aided Econometrics*. Taylor and Francis, New York.

Gilmore, L. A., Ravussin, E., Bray, G., Han, H., and Redman, L. (2014). An objective estimate of energy intake during weight gain using the intake-balance method. *American Journal of Clinical Nutrition*, 100:806–12.

Goldsmith, J., Liu, X., Jacobson, J., and Rundle, A. (2016). New insights into activity patterns in children, found using functional data analyses. *Journal of the American College of Sports Medicine*, pages 1723–9.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and and bayesian model determination. *Biometrika*, 82:711–32.

Hall, K. (2014). Estimating human energy intake using mathematical models. *American Journal of Clinical Nutrition*, 100:744–5.

Hall, K. D. (2010). Predicting metabolic adaptation, body weight change, and energy intrake in humans. *Am J Physical Endocrinol Metab*, 298:449–66.

Hall, K. D. and Chow, C. C. (2011). Estimating changes in free-living energy intake and its confidence interval. *American Journal of Clinical Nutrition*, 94:66–74.

Hand, G. A., Shook, R. P., Paluch, A. E., Baruth, M., Crowley, P., Jaggers, J. R., Prasad, V. K., Hurley, T. G., Hebert, J. R., O'Connor, D. P., Archer, E., Burgess, S., and Blair, S. N. (2013). The energy balance study: The design and baseline results for a longitudinal study of energy balance. *Research Quarterly for Exercise and Sport*, 84:275–86.

He, B., Bai, J. W., Zipunnikov, V., Koster, A., Caserotti, P., Lange-Maia, B., Glynn, N., Harris, T., and Crainiceanu, C. (2014). Predicting human movement with multiple accelerometers using movelets. *Journal of the American College of Sports Medicine*, pages 1859–66.

Hill, K., Dolmage, T., Woon, L., Goldstein, R., and Brooks, D. (2010). Measurement properties of the sensewear armband in adults with chronic obstructive pulmonary disease. *Thorax*, 65:486–91.

Hills, A., Mokhtar, N., and Byrne, N. (2014). Assessment of physical activity and energy expenditure: an overview of objective measures. *Frontiers in Nutrition*, 1:1–16.

Hind, K., Oldroyd, B., and Truscott, J. G. (2011). In vivo precision of the ge lunar idxa densitometer for the measurement of total body composition and fat distribution in adults. *European Journal of Clinical Nutrition*, 65:140–2.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15:199–236.

Huang, J.-H., Li, R.-H., Huang, S.-L., Sia, H.-K., Lee, S.-S., Wang, W.-H., and Tang, F.-C. (2017). Relationships between different types of physical activity and metabolic syndrome among taiwanese workers. *Scientific Reports*, 7:1–8.

Joe, H. and Zhu, R. (2005). Generalized poisson distribution: the property of mixture of poisson and comparison with negative binomial distribution. *Biometrical Journal*, 2:219–229.

Johannsen, D. and et al. (2010). Accuracy of armband monitors for measuring daily energy expenditure in healthy adults. *Medicine and Science in Sports and Exercise*, 42:2134–40.

Johnson, M. S. (2007). Modeling dichotomous item responses with free-knot splines. *Computational Statistics and Data Analysis*, 51:4178–92.

Jung-Min, L., Kim, Y., and Welk, G. J. (2014). Validity of consumer-based physical activity monitors. *Medicine and Science in Sports and Exercise*, 46:1840–8.

Kang, M., Rowe, D. A., Barreira, T. V., Robinson, T. S., and Mahar, M. T. (2009). Individual information-centered approach for handling physical activity missing data. *Research Quarterly for Exercise and Sport*, 80:131–7.

Kim, J., Tanabe, K., Yokoyama, N., Zempo, H., and Kuno, S. (2011). Association between physical activity and metabolic syndrome in middle-aged japanese: a cross sectional study. *BMC Public Health*, 11:624–31.

Kim, Y., Welk, Gregory, J., Braun, S. I., and Kang, M. (2015). Extracting objective estimates of sedentary behavior from accelerometer data: Measurement considerations for surveillance and research applications. *PLOS One*.

Kipnis, V., Freedman, L. S., Carroll, R. J., and Midthune, D. (2016). A bivariate measurement error model for semicontinuous and continuous variables: Application to nutritional epidemiology. *Biometrics*, 72:106–15.

Kipnis, V., Midthune, D., Buckman, D. W., Dodd, K. W., Guenther, P. M., Krebs-Smith, S. M., Subar, A. F., Tooze, J. A., Carroll, R. J., and Freedman, L. S. (2009). Modeling data with excess zeros and measurement error: Application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics*, 65:1003–10.

Lagerros, Y. and Lagiou, P. (2007). Assessment of physical activity and energy expenditure in epidemiological research of chronic diseases. *European Journal of Epidemiology*, 205:353–62.

Lee, J. A. and Gill, J. (2016). Missing value imputation for physical activity data measured by accelerometer. *Statistical methods in medical research*, pages 1–20.

Leitenstorfer, F. and Tutz, G. (2007). Generalized monotonic regression based on b-splines with an application to air pollution data. *Biostatistics*, 8:654–73.

Levin, J. A. (2005). Measurement of energy expenditure. *Public Health Nutrition*, 8:1123–32.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100:1989–2001.

Lindstrom, M. J. (2002). Bayesian estimation of free-knot splines using reversible jumps. *Computational Statistics and Data Analysis*, 41:255–69.

Liu, B., Yu, M., Braubard, B. I., Troiano, R. P., and Schenker, N. (2016). Mulitple imputation of completely missing repeated measures data within person from a complex sample: application to accelerometer data in the national health and nutrition examination survey. *Statistics in Medicine*, 35:5170–88.

Moore, J. X., Chaudhary, N., and Akinyemiju, T. (2017). Metabolic syndrome prevalence by race/ethnicity and sex in the united states, national health and nutrition examination survey. *Prev Chronic Dis*, 14.

Neelon, B., O'Malley, J. A., and Normand, S.-L. T. (2011). A bayesian two-part latent class model for longitudinal medical expenditure data: assessing the impact of mental health and substance abuse parity. *Biometrics*, 67:280–9.

Neelon, B., O'Malley, J. A., and Smith, V. A. (2016a). Modeling zero-modified count and semicontinuous data in health services research part 1: background and overview. *Statistics in Medicine*, 35:5070–93.

Neelon, B., O'Malley, J. A., and Smith, V. A. (2016b). Modeling zero-modified count and semicontinuous data in health services research part 2: case studies. *Statistics in Medicine*, 35:5094–112.

Neelon, B., Zhu, L., and Neelon, S. E. B. (2015). Bayesian two-part spatial models for semicontinuous data with application to emergency department expenditures. *Biostatistics*, 16:465–79.

Neilson, H. K., Robson, P. J., Friedenreich, C. M., and Csizmadi, I. (2008). Estimating activity energy expenditure: how valid are physical activity questionnaires? *American Journal of Clinical Nutrition*, 87:279–91.

Nusser, S., Carriquiry, A., Dodd, K., and Fuller, W. (1996). A semiparametric transformation approach to estimating usual daily intake distributions. *Journal of the American Statistical Association*, 91:1440–9.

ODPHP (2017). Office of disease prevention and health promotion-health.gov website. https://health.gov/paguidelines/guidelines/chapter2.aspx. Accessed: 2017-01-15.

Olsen, M. K. and Schafer, J. L. (2001). A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association*, 96:730–44.

Owen, N., Sparling, P., Healy, G., Dunstan, D., and Matthews, C. (2010). Sedentary behavior: Emerging evidence for a new health risk. *Mayo Clinic Proceedings*, 85:1138–41.

Papastamoulis, P. (2016). label.switching: An r package for dealing with the label switching problem in mcmc outputs. *Journal of Statistical Software*, 69:1–24.

Racette, S. B., Das, S. K., Bhapkar, M., Hadley, E. C., Robers, S. B., Ravussin, E., Pieper, C., DeLany, J. P., Kraus, W. E., Rochon, J., and Redman, L. M. (2011). Approaches for quantifying energy intake and %calorie restriction during calorie restriction interventions in humans: the multicenter calerie study. *Am J Physiol Endorinol Metab*, 302:441–8.

Reilly, J., Wilson, J., and Durnin, J. (1995). Determination of body composition from skinfold thickness: a validation study. *Archives of Disease in Childhood*, 73:305–10.

Reiner, M., Niermann, C., Jekauc, D., and Woll, A. (2013). Long-term health benefits of physical activity-a systematic review of longitudinal studies. *BMC Public Health*, 13.

Sanghvi, A., Redman, L. M., Martin, C. K., Ravussin, E., and Hall, K. D. (2015). Validation of an inexpensive and accurate mathematical method to measure long-term changes in free-living energy intake. *American Journal of Clinical Nutrition*, 102:353–8.

Santos-Lozano, A., Hernandez-Vicente, A., Perez-Isaac, R., Santin-Medeiros, F., Cristi-Montero, C., Casajus, J. A., and Garatachea, N. (2017). Is the sensewear armband accurate enough to quantify and estimate energy expenditure in healthy adults? *Annals of Translational Medicine*, 5.

Sarkar, A., Mallick, B. K., and Carroll, R. J. (2014a). Bayesian semiparametric regression in the presence of conditionally heteroscedastic measurement and regression errors. *Biometrics*, 70:823–34.

Sarkar, A., Mallick, B. K., Staudenmayer, J., Pati, D., and Carroll, R. J. (2014b). Bayesian semi-parametric density deconvolution in the presence of conditionally heteroscedastic measurement errors. *Journal of Computational and Graphical Statistics*, 23:1101–25.

Scheers, T., Philippaerts, R., and Lefevre, J. (2013). Compliance with different physical activity recommendations and its association with socio-demographic characteristics using an objective measure. *BMC Public Health*, 13:1–10.

Schrack, J., Zipunnikov, V., Goldsmith, J., Bai, J., Simonsick, E., Crainiceanu, C., and Ferrucci, L. (2014). Assessing the "physical cliff": Detailed quantification of age-related differences in daily patterns of physical activity. *Journals of Gerontology*, 69:973–9.

Scollnik, D. P. (1998). On the analysis of the truncated generalized poisson distribution using a bayesian method. *Astin Bulletin*, 28:135–152.

Sinha, S., Mallick, B. K., Kipnis, V., and Carroll, R. J. (2010). Semiparametric bayesian analysis of nutritional epidemiology data in the presence of measurement error. *Biometrics*, 66:444–54.

Sisson, S. B., Camhi, S. M., Church, T. S., Tudor-Locke, C., Johnson, W. D., and Katzmarzyk, P. T. (2010). Accelerometer-determined steps/day and metabolic syndrome. *American Journal of Preventive Medicine*, 38:575–82.

Smith, S. A., Ansa, B. E., Yoo, W., Whitehead, M. S., and Coughlin, S. S. (2016). Determinants of adherence to physical activity guidelines among overweight and obese african american breast cancer survivors: implications for an intervention approach. *Ethnicity and Health*, pages 1–13.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *JRSS Series B*, 64:583–639.

Steeves, J., Murphy, R., Crainiceanu, C., Zipunnikov, V., Glass, T., Van Domelen, D., and Harris, T. (2015a). Daily patterns of physical activity by type 2 diabetes definition: Comparing diabetes, prediabetes, and participants with normal glucose levels in nhanes 2003-2006. *Preventive Medicine Reports*, 2:152–7.

Steeves, J., Murphy, R., Zipunnikov, V., Strath, S., and Harris, T. (2015b). Women workers and women at home are equally inactive: Nhanes 2003-2006. *Journal of the American College of Sports Medicine*, pages 1635–42.

Stephens, M. (2000). Dealing with label switching in mixture models. *JRSS Series B*, 62:795–809.

Takezawa, K. (2006). *Introduction to Nonparametric Regression.* Wiley Series in Probability and Statistics.

Thomas, D. M., Bouchard, C., Church, T., Slentz, C., Kraus, W., L.M., R., Martin, C., Silva, A., Vossen, M., Westererp, K., and Heymsfield, S. (2012). Why do individuals not lose more weight from an exercise intervention at a defined dose? an energy balance analysis. *Obesity Reviews.*

Thomas, D. M., Martin, C. K., Heymsfield, S., Redman, L. M., Schoeller, D. A., and Levine, J. A. (2011a). A simple model predicting individual weight change in humans. *Journal of Biological Dynamics*, 5:579–599.

Thomas, D. M., Martin, C. K., Heymsfield, S., Redman, L. M., Schoeller, D. A., and Levine, J. A. (2011b). A simple model predicting individual weight change in humans. *Journal of Biological Dynamics*, 5:579–99.

Thomas, D. M., Schoeller, D. A., Redman, L. A., Martin, C. M., Levine, J. A., and Heymsfield, S. B. (2010). A computational model to determine energy intake during weight loss. *American Journal of Clinical Nutrition*, 92:1326–31.

Tooze, Janet, A., Grunwald, G. K., and Jones, R. H. (2002). Analysis of repeated measures data with clumping at zero. *Statistical Methods in Medical Research*, 11:341–55.

Tooze, J. A., Midthune, D., Dodd, K. W., Freedman, L. S., Krebs-Smith, S. M., Subar, A. F., Guenther, P. M., Carroll, R. J., and Kipnis, V. (2006). A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *Journal of the American Dietetic Association*, 106:1575–87.

Tucker, J. M., Welk, G. J., and Beyler, N. K. (2011). Physical activity in u.s. adults compliance with the physical activity guidelines for americans. *American Journal of Preventive Medicine*, 40:454–61.

Tucker, J. M., Welk, G. J., Beyler, N. K., and Kim, Y. (2016). Associations between physical activity and metabolic syndrome: comparison between self-report and accelerometry. *American Journal of Health Promotion*, 30:155–62.

Urbanek, J., Harezlak, J., Glynn, N., Harris, T., Crainiceanu, C., and Zipunnikov, V. (2017). Stride variability measures derived from wrist- and hip-worn accelerometers. *Gait and Posture*, 52:217–23.

Wang, W. and Small, D. S. (2015). Monotone b-spline smoothing for a generalized linear model response. *The American Statistician*, 69:28–33.

Warburton, D. E., Nicol, C. W., and Bredin, Shannon, S. (2006). Health benefits of physical activity:the evidence. *Canadian Medical Association Journal*, 176:801–9.

Welk, G., Schaben, J., and Morrow, J. J. (2004). Reliability of accelerometry-based activity monitors: a generalizability study. *Med Sci Sports Exerc*, 36:1637–45.

Wells, J. C., Fuller, N. J., Dewit, O., Fewtrell, M. S., Elia, M., and Cole, T. J. (1999). Four-component model of body composition in children: density and hydration of fat-free mass and comparison with simpler models. *American Journal of Clinical Nutrition*, 69:904–12.

Xiao, L., Huang, L., Schrack, J., Zipunnikov, V., and Crainiceanu, C. (2015). Quantifying the lifetime circadian rhythm of physical activity: a covariate-dependent functional approach. *Biostatistics*, 16:352–67.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57:348–68.

Zellner, A. and Ando, T. (2010). Bayesian and non-bayesian analysis of the seemingly unrelated regression model with student-t errors, and its application for forecasting. *International Journal of Forecasting*, 26:413–434.

Zhang, S., Krebs-Smith, S. M., Midthune, D., Perez, A., Buckman, D. W., Kipnis, V., Freedman, L. S., Dodd, K. W., and Carroll, R. J. (2011a). Fitting a bivariate measurement error model for episodically consumed dietary components. *The International Journal of Biostatistics*, 7:1–31.

Zhang, S., Midthune, D., Guenther, P. M., Krebs-Smith, S. M., Kipnis, V., Dodd, K. W., Buckman, D. W., Tooze, J. A., Freedman, L. S., and Carroll, R. J. (2011b). A new multivariate measurement error model with zero-inflated dietary data, and its application to dietary assessment. *The Annals of Applied Statistics*, 5:1456–87.

Zhao, G., Ford, E., Li, C., and Mokdad, A. (2008). Compliance with physical activity recommendations in us adults with diabetes. *Diabetic Medicine*, 25:221–7.

APPENDIX. ADDITIONAL MATERIAL

A.1 Full conditional distributions for the parameters in Chapter 3

$$\gamma|\cdot \propto \left[\prod_{i=1}^{1057}\prod_{j=1}^{2} f(Y_{1ij}|Z_i, b_{1i}, \gamma, \lambda) f(Y_{2ij}|Z_i, b_{2i}, \beta, \sigma_y^2, \pi_i)\right] p(\gamma) \tag{1}$$

$$= \left[\prod_{i=1}^{1057}\prod_{j=1}^{2} \mu_1(1-\lambda)(\mu_1(1-\lambda)+Y_{1ij}\lambda)^{Y_{1ij}-1} e^{-\mu_1(1-\lambda)-Y_{1ij}\lambda} \frac{1}{Y_{2ij}\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2\sigma_y^2}(lnY_{2ij}-Z_i`\beta-b_{2i})^2}\right] \tag{2}$$

$$\times e^{-\frac{1}{200}\gamma^2} \tag{3}$$

$$\text{where } \mu_1 = e^{Z'\gamma+b_{1i}} \tag{4}$$

$$\beta|\cdot \propto \left[\prod_{i=1}^{1057}\prod_{j=1}^{2} f(Y_{2ij}|Z_i, b_{2i}, \beta, \sigma_y^2, \pi_i) I(Y_{2ij}>0)\right] p(\beta) \tag{5}$$

$$\sim N(m_\beta, V_\beta) \tag{6}$$

$$V_\beta = (V_0^{-1} + \frac{1}{\sigma_y^2} Z'Z)^{-1}, m_\beta = V_\beta(V_0^{-1} m_0 + Z'(log\mathbf{Y_2} - \mathbf{b_2})/\sigma_y^2), \text{ for } Y_{2ij}>0 \tag{7}$$

$$\lambda|\cdot \propto \left[\prod_{i=1}^{1057}\prod_{j=1}^{2} f(Y_{1ij}|Z_i, b_{1i}, \gamma, \lambda) f(Y_{2ij}|Z_i, b_{2i}, \beta, \sigma_y^2, \pi_i)\right] p(\lambda) \tag{8}$$

$$= \left[\prod_{i=1}^{1057}\prod_{j=1}^{2} \mu_1(1-\lambda)(\mu_1(1-\lambda)+Y_{1ij}\lambda)^{Y_{1ij}-1} e^{-\mu_1(1-\lambda)-Y_{1ij}\lambda} \frac{1}{Y_{2ij}\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2\sigma_y^2}(lnY_{2ij}-Z_i`\beta-b_{2i})^2}\right] \tag{9}$$

$$\times \lambda^{a_\lambda-1} e^{-\frac{\lambda}{b_\lambda}} \tag{10}$$

$$\sigma_y^2|\cdot \propto \left[\prod_{i=1}^{1057}\prod_{j=1}^{2} f(Y_{2ij}|Z_i, b_{2i}, \beta, \sigma_y^2, \pi_i) I(Y_{2ij} > 0)\right] p(\sigma_y^2) \tag{11}$$

$$\sim \text{Inverse-Gamma}\left(\frac{N^*}{2} + a_0, \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{2}(logY_{2ij} - Z_i'\beta - b_{2i})^2\right) \tag{12}$$

$$\text{for } Y_{2ij} > 0 \text{ where } N^* \text{ is the number of non-zero observations of } Y_{2ij} \tag{13}$$

$$\Sigma_b|\cdot \propto \left[\prod_{i=1}^{n} f(b_{1i}, b_{2i}|\Sigma_b)\right] p(\Sigma_b) \tag{14}$$

$$\sim \text{Inverse-Wishart}(n + d_0, \mathbf{b'b} + D_0) \tag{15}$$

$$\text{where } \mathbf{b} \text{ is an } n \times 2 \text{ matrix with the first column containing elements of } b_1, \tag{16}$$

$$\text{and second column containing elements of } b_2 \tag{17}$$

$$b_{1i}, b_{2i}|\cdot \propto \left[\prod_{j=1}^{2} f(Y_{1ij}|Z_i, b_{1i}, \gamma, \lambda) f(Y_{2ij}|Z_i, b_{2i}, \beta, \sigma_y^2, \pi_i)\right] f(b_{1i}, b_{2i}|\Sigma_b) \tag{18}$$

$$= \left[\prod_{j=1}^{2} \mu_1(1-\lambda)(\mu_1(1-\lambda) + Y_{1ij}\lambda)^{Y_{1ij}-1} e^{-\mu_1(1-\lambda)-Y_{1ij}\lambda} \frac{1}{Y_{2ij}\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2\sigma_y^2}(lnY_{2ij} - Z_i`\beta - b_2 i)^2}\right] \tag{19}$$

$$\times e^{-\frac{1}{2}\mathbf{b_i}'\Sigma_b^{-1}\mathbf{b_i}} \tag{20}$$

A.2 Full conditional distributions for the parameters in Chapter 4

$$P(\zeta_i = k)|\cdot \propto f(Y_i|\zeta_i, \theta) f(\zeta_i|\mathbf{p}) \propto p_k N(\lambda_k + m(t_i, \gamma), \Sigma_k) \tag{21}$$

$$\mathbf{p}|\cdot \propto f(\zeta_i|\mathbf{p}) f(\mathbf{p}) = \text{Cat}(H, \mathbf{p})\text{Dirichlet}(a) \sim \text{Dirichlet}(a + \tilde{n}) \tag{22}$$

where (23)

$$\tilde{n} = (n_1, ..., n_H) \tag{24}$$

$$n_h = \sum_{i=1}^{n} I(\zeta_i = h) \tag{25}$$

$$\lambda_k|\cdot \propto \prod_{i=1}^{n} f(Y_i|\zeta_i = k, \lambda_k, \gamma, \Sigma_k) f(\lambda_k) \sim N(m_{\lambda_k}, V_{\lambda_k}), \forall k \tag{26}$$

where (27)

$$V_{\lambda_k} = \left(V_0^{-1} + n_k \Sigma_k^{-1}\right)^{-1} \tag{28}$$

$$m_{\lambda_k} = V_{\lambda_k} \left(V_0^{-1} m_0 + \Sigma_k^{-1} (\mathbf{Y^{(k)}} - \mathbf{m}(\mathbf{t}, \gamma))\right) \tag{29}$$

$\mathbf{Y^{(k)}}$ is response for individuals in component k (30)

$$\Sigma_k|\cdot \propto f(Y_i|\zeta_i = k, \lambda_k, \gamma, \Sigma_k) f(\Sigma_k) \sim \text{Inverse-Wishart}(d_0 + n_k, D_0 + D) \tag{31}$$

$$D = (\mathbf{Y^{(k)}} - \lambda_{\mathbf{k}} - \mathbf{m}(\mathbf{t}, \gamma))'(\mathbf{Y^{(k)}} - \lambda_{\mathbf{k}} - \mathbf{m}(\mathbf{t}, \gamma)) \tag{32}$$

$$\gamma|\cdot \propto \prod_{i=1}^{n} f(\mathbf{Y_i}|\zeta, \lambda_k, \gamma, \Sigma_k) f(\gamma) \tag{33}$$

$$\propto \prod_{i=1}^{n} e^{\frac{1}{2}(\mathbf{Y_i} - \lambda_{\mathbf{h(i)}} - \mathbf{m}(\mathbf{t_i}, \gamma))' \Sigma_{h(i)}^{-1} (\mathbf{Y_i} - \lambda_{\mathbf{h(i)}} - \mathbf{m}(\mathbf{t_i}, \gamma))} e^{\frac{1}{2}(\gamma - m_\gamma)' V_\gamma^{-1} (\gamma - m_\gamma)} \tag{34}$$

where $h(i)$ is an indicator which component Y_i belongs to (35)